EUROPEAN JOURNAL
FOR PHILOSOPHY OF RELIGION

VOLUME 6 NUMBER 4 WINTER 2014

ARTICLES

BOOK REVIEWS AND NOTICES

COMMUNAL AND INSTITUTIONAL TRUST: AUTHORITY IN RELIGION AND POLITICS

C.A.J. COADY

University of Melbourne

Linda Zagzebski's book on epistemic authority is an impressive and stimulating treatment of an important topic.[1] I admire the way she manages to combine imagination, originality and argumentative control. Her work has the further considerable merit of bringing analytic thinking and abstract theory to bear upon areas of concrete human concern, such as the attitudes one should have towards moral and religious authority. The book is stimulating in a way good philosophy should be – provoking both disagreement and emulation.

I agree with much of what she says, and have been instructed by it, but it will be of more interest and relevance here if I concentrate upon areas of disagreement. Perhaps they are better seen as areas, at least some of them, where her emphases suggest a position that seems to me untenable, but that she may not really intend. In that event, I will be happy to have provoked a clarification or the dispelling of my misunderstanding.

My focus will be upon problems in her account of communal authority and autonomy, especially with respect to religious and political authority. Here my worry is that she places too much trust in trust and not enough in what I call selective mistrust.

THE EPISTEMOLOGY OF TESTIMONY: SOME PRELIMINARIES

I have in the past written quite a bit on a topic central to Zagzebki's discussion, namely the role of testimony in our intellectual life, especially in my book *Testimony: a Philosophical Inquiry.*[2] Moreover, the relation of

[1] Linda Trinkaus Zagzebski, *Epistemic Authority: A Theory of Trust, Authority, and Autonomy in Belief* (Oxford: Oxford University Press, 2012). References to this book will be bracketed in the text.

[2] C.A.J. Coady, *Testimony*: A *Philosophical Study* (Oxford: Clarendon Press, 1992).

testimonial authority to autonomy was something I visited a little more recently in a paper on 'Testimony and Intellectual Autonomy'.[3] I think that much of what I said there is congruent with Zagzebski's discussion of the way an individual's autonomy can be consistent with a certain deference to the authority of another's testimony and even of their beliefs revealed in non-testimonial fashion, but I detect a dissonance (to use one of her favoured expressions) between us when it comes to the authority of religious and political communities.

Where we agree is in rejecting what she calls epistemic egoism; a version of which embodies what J. L. Mackie once discussed as an ideal of autonomous knowledge built into the empiricist tradition, namely, the view that our wide-ranging dependence on testimony could only be admitted as knowledge or reliable belief if (to quote him) 'the knower somehow checks for himself the credibility of the witness' whenever he relies upon one.[4] It is clear from the rest of the Mackie passage that 'checks for himself' means relying solely upon checks that use only the knower's individual resources and are quite independent of anybody's testimony. There are actually two possible interpretations of the egoism or individualism inherent in the tradition as Zagzebski makes clear in her book.

The first (what she calls 'extreme egoism') admits no role for the transfer of knowledge from others and restricts one's knowledge to what one can acquire solely by one's own individual epistemic abilities. This would, of course, lead to an extraordinarily narrow knowledge base for any person, though it is possibly what Plato had in mind in his comments on how testimony could not provide the logos that true knowledge (or understanding) requires. Interestingly, as Zagzebski notes, Elizabeth Fricker refers to this extreme position as 'intellectual autonomy', but what Zagzebski calls 'moderate' or 'standard' egoism and what I have called reductionism is more in tune with the tradition Mackie articulates, certainly in David Hume's discussion of reports of miracles.[5] The extreme form is however still in play when people discuss moral autonomy since the idea that, in moral matters you must reach

[3] C. A. J. Coady, 'Testimony and intellectual autonomy', *Studies in History and Philosophy of Science*, Part A 33 (2):355-372 (2002).

[4] J. L. Mackie, 'The Possibility of Innate Knowledge', *Proceedings of the Aristotelian Society*, 70:245-257 (1969), p. 254.

[5] David Hume, *An Enquiry Concerning Human Understanding*, section 10 'On Miracles', 2nd edition (Oxford: Oxford University Press, 1957).

moral knowledge wholly by reliance upon your own resources has much more popularity. Zagzebski argues, correctly in my view, that there are serious problems with it, but I will not discuss moral knowledge directly.

I argued in my book that such egoism or reductionism about testimony was doomed to failure since it didn't take seriously the depth of our reliance upon the word of others. I criticised attempts to vindicate that reliance by recourse alone to individual sources of knowledge or reliable belief such individual perceptions, memories and inferences as doomed to failure. Not only were such attempts vitiated by impracticality (as H. H. Price had argued) because of the time-consuming amount of individual checking on the reliability of witnesses or other types of testifiers that the posture entailed, but most of the procedures that were apparently reliant only upon my unaided powers of observation, memory and inference were actually infected at core by further reliance upon unchecked interpersonal sources. The concepts of 'observation' and 'experience' that figured in such individualist efforts were invariably, at least in part, appeals to common experiences and observations rather than the individual's alone. Moreover, there were profound difficulties in determining the correlations between types of report and types of situation that the project required without again begging the question. Nor could the project be restricted to determining the reliability of this or that witness on this or that topic since what was required was a more general justification of, as it were, the institution of testimony. There is much more that could be said of this but I cannot say it here.

It is worth noting, however, that the widespread neglect of the topic of testimony that I originally complained of has been dramatically remedied and a debate between reductionist and non-reductionist theories has developed apace with much sophistication and complex distinctions on both sides. I do not intend entering this debate directly here, but I am unpersuaded that the egoist position has been restored by that debate.[6]

[6] Some important contributions in book form are: Jonathan E. Adler, *Belief's Own Ethics* (Cambridge MA: The MIT Press, 2002); Jennifer Lackey, *Learning From Words: Testimony as a Source of Knowledge* (Oxford: Oxford University Press, 2008); Jennifer Lackey and Ernest Sosa (eds), *The Epistemology of Testimony* (Oxford: Oxford University Press, 2006); and Paul Faulkner, *Knowledge on Trust* (Oxford: Oxford University Press, 2011). There are many other significant journal articles on the topic including Elizabeth Fricker, 'Telling and Trusting: Reductionism and Anti-Reductionism in the Epistemology of Testimony', *Mind* 104 (414):393-411 (1995); and Peter J. Graham, 'Testimonial Justification: Inferential or Non-Inferential?', *The Philosophical Quarterly*, 56 (2006), 84–95.

The debate has encompassed other dimensions concerned with the moral or evaluative aspects associated with the attitudes of trust, acceptance, and mode of delivery involved in giving and receiving testimony. These have fed into a parallel philosophical movement that began in the 1980s investigating trust as a phenomenon and attitude in human affairs. The moral and epistemological aspects of trusting have been discussed by some excellent philosophers, many of them women, and most notably by Annette Baier.[7] The epistemological aspect of this complex of investigations has often developed under the heading of social epistemology.

ZAGZEBSKI'S POSITIVE ACCOUNT AND THE ROLE OF SELF-TRUST

One issue that is nested in these discussions is that of intellectual or epistemic autonomy and its relation to authority. Zagzebski's account of this relation requires a version of anti-reductionism though her theory goes beyond reliance on testimony. Her theory is not content with showing the futility of the reductionist position, but argues for a non-egoist justification of our epistemic reliance on others and hence on the authority they have for us; this argument makes a positive virtue of beginning from the first-person perspective and from the idea of trust in the self. This strategy is interesting and original precisely because it recognizes some of the strength in the egoist's insistence on the importance of the intellectual standing of the self. There are two key ideas here. The first is that an individual's own epistemic base, the base that is to underpin epistemic autonomy, and is relied upon by the epistemic egoist, rests on nothing other than a form of trust – trust in the self's cognitive powers. So the very powers of personal observation, memory and inference are themselves inevitably taken on trust. This conclusion arises from the fact that there is a circularity in any attempt to justify an individual's basic epistemic resources since ultimately any such attempt relies upon those same resources. Many epistemologists, such as Richard Foley and William Alston, have recognised this, sometimes through consideration of the challenge of extreme scepticism and sometimes simply by reflecting upon the nature of rationality. Whether this resort 'refutes' scepticism or not, it is clear that it renders the attempt to use our reasoning powers

[7] Annette Baier has made numerous contributions to the trust discussion beginning with her influential article, 'Trust and Antitrust' in *Ethics, Ethics* 96 (2):231-260 (1986).

to reject our reasoning powers curiously pointless for any endeavour to sketch the lineaments of our epistemological landscape. But Zagzebski goes further than this in arguing that the trust is not only forced upon us by the circularity, it is a requirement of another feature of our nature, or even a requirement of rationality, namely the need to reduce or eliminate dissonance. This dissonance operates at two levels. Since she thinks of rationality as 'doing a better job of what we do naturally' (p. 45) it requires self-reflection upon our natural inclination and apparent powers to seek truth. Awareness of the circularity creates a dissonance for the self-reflective person between trusting her natural epistemic faculties and believing that they are untrustworthy. It also creates a dissonance between the feeling that one's faculties are trustworthy and the feeling that they are not. The self-reflective person will need to resolve this dissonance as well. Compare the person who has every reason to believe that she has locked her house but has an obsessive doubt about it and not only returns home several times to check on it, but constantly feels that she ought through the rest of the day.[8] It is possible to persist in such behaviour yet it is not only disabling, but clearly subject to the charge of irrationality. Such a charge is even more plausible where the dissonance is across the board of all one's natural epistemic powers. Hence, for Zagzebski, self-trust is a requirement of rationality. Further to this she argues that although self-reflection yields a rational trust in our faculties, it requires conscientious self-reflection in the use of those faculties to have a good chance of getting at the truth. As she puts it:

> A conscientious person has evidence she is more likely to get the truth when she is conscientious, but she trusts evidence in virtue of her trust in herself when she is conscientious, not conversely. Her trust in herself is more basic than her trust in evidence, and that includes evidence of reliability. The identification of evidence, the identification of the way to handle and evaluate evidence, and the resolution of conflicting evidence all depend on the more basic property of epistemic conscientiousness. (p. 49)

Her next move is to generalise the trust involved to encompass trust in others. This requires a sort of universalising principle that has a Kantian flavour. Since I find that others have broadly the same epistemic capacities that I trust in myself I have no reason not to treat their situation as similar

[8] This is a variation on an example of my colleague Karen Jones, cited by Zagzebski in the text.

to mine with respect to self-trust and the exercise of conscientious reflection. So she thinks that consistency demands that 'I have the same basic trust in the epistemic faculties of all other persons whose general similarity to me I come to believe when I am conscientious' (p. 160). She thinks there is a prima facie case for the same conclusion with respect to moral beliefs. This leads to what she calls a weak form of epistemic universalism which she expresses as 'the fact that another person has a certain belief always gives me prima facie reason to believe it' (p. 58).

SOME PRINCIPLES AND THEIR IMPLICATIONS

Out of a great deal of close discussion of these matters she develops a variety of principles concerning more stringent justification of beliefs and emotions gained from others. I cannot explore all of this but a flavour of the principles in question can be gained from mentioning a few of them. First is one that makes a connection, as do several others, with the idea of authority and indirectly with that of autonomy. Consider:

> **Justification Thesis 2 for the Authority of Belief (JAB 2)**
> The authority of another person's belief for me is justified by my conscientious judgment that I am more likely to form a belief that survives my conscientious self-reflection if I believe what the authority believes than if I try to figure out what to believe myself.

It is noteworthy that this principle embraces more than justification of the testimony of others since what they believe may be shown in behaviour rather than explicit testimony. It is thus rather stronger than a testimony principle and perhaps more debatable since someone who seriously tells us that p is vouching for it in a way that their non-testimony behavioural manifestations of belief typically do not. This difference is important and is connected to Zagzebski's derivation of other-trust from self-trust, but I shall merely note it here, and will concentrate my criticisms mostly on areas where epistemic trust is in play in dealing with explicit religious (and political) statements of belief.[9]

[9] Her principle concerning trust in testimony is formally parallel to this trust in belief principle (though actually she has two formulations of both) and goes as follows:

'Justification Thesis 2 for the Authority of Testimony (JAT 2)

The authority of another person's testimony for me is justified by my conscientious judgment that if I believe what the authority tells me, the result will survive my conscientious self-reflection better than if I try to figure out what to believe myself.' (p. 133)

It is also slightly puzzling that principle JAB 2 begins with phrase 'the authority of another person's belief' and concludes with a reference to that person's belief as 'what the authority believes'. This is puzzling as that person is not introduced as 'an authority' but simply as 'another person'. Since she thinks that any other person has some prima facie authority for us in their beliefs this may be all she means by calling any 'other person' an authority. But this is not the usual sense of 'an authority', nor is it quite the sense that she later relies upon. This is important for my later discussion of communal belief and communal authority.

Later she provides a community version of the principle that goes as follows:

> **Justification of Communal Epistemic Authority 2 (JCEA2)**
> The authority of my community is justified for me by my conscientious judgment that if I believe what We believe, the result will survive my conscientious self-reflection better than if I try to figure out what to believe in a way that is independent of US.

Several points of clarification are needed here. The references to 'we' and 'us' in the communal authority case is partly based on the common use of these pronouns to refer to communities to which individuals belong, but she takes this use to be a good indicator that the communities we belong to are what she calls 'extended selves'. She recognises that the forms of authority in such communities differ widely in those that have democratic structures to determine authority and those that are more top-down. She also recognises the fallibility of the extended self but argues that its imperfections are no barrier to the judgement that we do better at getting the truth by relying on the authority than by going it alone. If we make the conscientious judgement that the authority's record is so bad that we do better making the judgements without it then we need to transfer to another community or build a new community (p. 158). That we can reasonably reject an authority's belief (or testimony) is admitted by Zagzebski early in her discussion because it is a general trust in ourselves that leads to our trust in authority and that self-trust may be exercised to reject the authority's belief, if for example, I conscientiously judge that it is clearly inconsistent with other conscientiously determined beliefs.

She carries this apparatus of arguments and principles over to the discussion of moral and religious epistemic authority and touches upon political authority. In this endeavour she is committed to there being

truth attainable in moral and religious matters and she argues for this commitment. Some, perhaps many, will dispute this, but I am going to accept it, partly because I believe it to be true, but mostly because I want to see whether what she makes of it in terms of authority is cogent.

Another of Zagzebski's important discussions concerns the role of emotions and of exemplars in the acquisition of knowledge. I cannot explore this fully here, but mention it at this point because of its relevance to some of her claims about religious, moral and political authority that I will criticise later. She holds that a conscientious person should trust her emotional faculties in the same way that she should trust her cognitive faculties since although there is no non-circular justification for the reliability of either yet the 'outputs of both can survive conscientious self-reflection' (p. 86). Of course some emotions do not match appropriately the circumstances in our lives but others can be judged appropriate by a self-conscious attempt to fit them to their objects. Amongst these emotions that can be trusted are indignation, sympathy, the feel for the ridiculous, and admiration. Admiration is significant in leading to the idea of exemplars. We can admire qualities and acts but also persons who display them, and we can and should learn from them. Zagzebski argues that this applies across the board to the intellectual, moral and practical qualities that are encapsulated in the concept of wisdom, until recently little explored by philosophers (pp. 81ff.). This learning from exemplars often involves a role for imitation.

COMMUNAL AUTHORITY IN RELIGION AND IN POLITICS

This account of Zagzebski's position is inevitably sketchy and oversimplified. I can only hope that it gives a fair understanding of her main line of thought sufficient to raise some issues about her account of intellectual autonomy and communal authority in religious, moral and political matters. In order to address that I will pass over difficulties raised by her argument from dissonance where, for example, one might question the strength of the need to resolve dissonance on which she relies so much, or the precise details of the consistency or universalisability argument she uses in the foundation of her theory, and so on. I think she can probably be defended against objections along these lines but I will not attempt that here. I will however say something about dissonance and the self in relation to trust in institutional authority.

Her defence of authority in the epistemic arena is anti-reductionist in a way that, as we have seen earlier, runs counter to certain common ideological outlooks about authority and autonomy. I agree with her idea that, to put it in different terms, intellectual authority is not an external, as it were, regrettable necessity to be treated invariably with suspicion, but I think that her treatment of communal authority is unsatisfactory in certain respects. The idea of a community as an extended self has some obvious merits but the extension is rarely as straightforward as her treatment often suggests. The discourse of 'we' and 'our' is real enough, but the referent of these expressions is rarely straightforward. The nuances in such references are particularly complex in the case of political authority to which she does not pay a lot of attention. A political community is invariably composed of sub-communities that are either directly political themselves, as with political parties, or communities devoted in the first place to ends other than the political but with interests, values and beliefs that bear upon the political and drive political thinking. So an individual's extended self is likely to be extended in several different, potentially clashing directions. A person who sees herself as a member of the mining community, the Anglican church, the aboriginal community, and her small township is likely to have respected exemplars in each of these 'extended selves' and will often have to do some strong negotiating to avoid a schizophrenic 'extended self'. Zagzebski links her discussion of epistemic authority to the project that an 'executive self' has of harmonising her desires, emotions and beliefs in pursuit of fulfilment, and there is much to be said for this. But the executive self who self-reflectively examines this situation will have to come to terms with clashes *internal* to her community. This will often enough involve admiration for different authority figures with different key beliefs amongst some or many overlapping beliefs. Some of these will be contemporary figures and others figures from the near or distant past. Before looking further at politics, I want to examine what she says about religious communities with the above discussion as background because, in addition to the intrinsic interest of religious authority (and autonomy), there are certain parallels between religious and political communities.

A cautionary note is that Zagzebski, in discussing the shortcomings of Joseph Raz's definition of political authority as a model for authority more generally, concentrates on small communities which suggests that she has some misgiving about extending the account she develops

regarding them to larger political communities. It is I think more likely that she focuses on small communities to show more readily her differences with Raz rather than as a sign that she cannot extend her account more broadly, since she finishes with principles of justification of communal authority that contain no such qualification about small size, and in any case small communities are often composed of smaller communities again with similar problems to those I have hinted at. Moreover, she later applies her principles to huge communities like the Catholic Church, so it seems size is not a reasonable differentiating factor. She says some other things about Raz and political authority that invoke something different and I will return to that.

When Zagzebski discusses religious epistemic authority she notes that different religions have evolved very different structures or techniques to convey or exercise communal authority and these range from highly formal edifices like the 'teaching office of the Catholic Church' (p. 176) to the informal immersion in a way of life like the Old Order Amish. I think the contrast here is not as sharp as it appears, but the Catholic community, tradition and teaching authority is in any case understood by Zagzebski in too monolithic a fashion, admittedly a fashion encouraged by the institution's formal leadership. Within any existing community of knowledge or belief, including religions, there will be various sub-communities sharing some respect for the authority of other sub-communities, but differing from them in important ways, and similarly there will be different exemplars favoured by different members of such sub-communities. Think, for instance, of the philosophy community within a country or even within a particular university.

In speaking of the way that we accept or reject communal authority, Zagzebski appears, at least some of the time, to think of it as an all or nothing affair in what seems to me an unrealistic way. She admits, as we saw earlier, that if we make the conscientious judgement that the authority's record is so bad that we do better making the judgements without it then we need to transfer to another community or build a new community (p. 158). But an autonomous self in a religious community negotiates its membership and its shared beliefs in a much more dynamic and interactive way than this picture suggests. To take the example of the Catholic Church: its highly authoritative structure and the edicts of its official leadership have often in the past been at odds with widespread beliefs of sections of the faithful and continue (even more dramatically) to be so today. The great degree of this rift is evident in both doctrinal,

moral and 'disciplinary' matters as is evidenced by a variety of polling figures of ordinary Catholics' beliefs about the morality of contraception, homosexuality, divorce, abortion, the doctrines of hell and 'no salvation outside the Church', and the disciplines of clerical celibacy and an exclusively male priesthood.[10] Gallup Poll figures in 2009 showed that 40 percent of Catholics in the USA (compared to 41% of non-Catholics) found abortion 'morally acceptable'.[11] Nor can this sort of result be put down to simple lay ignorance or backsliding since the rift is also present amongst regular lay churchgoers and similar disagreement is evident amongst clergy and theologians. For instance, even among those who attend church once a week or more, 83 percent of sexually active Catholic women use a form of contraception banned by the Vatican.[12] Such divergence has sometimes led to splintering and the formation of new communities or departure from religious community altogether, and that fits Zagzebski's picture. But it has also been contained within the community so that quite different positions on what it is necessary to believe have been maintained without splitting the 'We' and 'Us' extended selves.[13] How can this be?

One way of dealing with this is to distinguish between core and peripheral beliefs, and argue that as long as the core beliefs are centrally authoritative for the community members, there is room for diversity on the others. This distinction has some merit, but determining the division between what is core and what is peripheral is often enough something

[10] Just some of the recent evidence from respectable polls can be found (for the United States) in regular reports in the journal *Catholics for Choice*. For a summary of findings see: *Catholics for Choice. (2011). The Facts Tell the Story*, Catholics and Choice, Washington DC: <http://www.catholicsforchoice.org/documents/Factstellthestoryweb.pdf> [accessed 30/01/2014]. Recently, in the UK, the sociologist, Professor Linda Woodhead, conducted a poll (administered by YouGov) that reported similar findings. See *http://faithdebates.org.uk/research/*. For Woodhead's own summary of the research see her essay, 'New Poll: "Faithful Catholics" an Endangered Species', in *Religion Dispatches*, January 20, 2014. Available at: <http://www.religiondispatches.org/archive/culture> [accessed 25/11/2014].

[11] See: <http://www.gallup.com/poll/117154/Catholics-Similar-Mainstream-Abortion-Stem-Cells.aspx> [accessed 30/01/2014].

[12] *Catholics for Choice*, 2011.

[13] An interesting example of tolerated dissent within the wider Catholic community on an important issue is the fierce and unresolved dispute between the Jesuits and Dominicans in the 16th and 17th centuries about the nature of grace and its relation to free will. This was an issue at the heart of theological controversy leading to the Catholic/Protestant split in the Reformation, so its failure to disrupt Catholic unity in this later period is instructive.

that authority will want a say upon, and it is often enough, precisely that 'say' which is contested. Moreover, 'peripheral' need not mean unimportant; it may signal merely that the issue is not one on which authenticity of one's membership in the community should turn. In the case of the Catholic Church, for instance, it seems to me as a Catholic (and of course to others, Catholic or not) that the ordination of women priests is such a peripheral matter, as is the celibacy of the clergy. Yet the formal 'teaching office' sees things differently.

THE COMPLEXITIES OF INSTITUTIONAL AUTHORITY

This raises the question of institutional authority. Communities seem inevitably to generate their distinctive institutions, and these equally inevitably make claims to a certain authority both practical and epistemic. The practical dimension can be more or less coercive, but always involves power of some sort. That is very characteristic of politics, but exists elsewhere.

The question of power returns us to the ambiguity about the nature of the authority of others that I noted earlier in Zagzebski's expression of Principle JAB 2. The fact that each person has some presumptive authority for others about their beliefs (if it is a fact) is different from the authority that designated 'authorities' have for their beliefs in certain areas. We can assume that ordinary folk usually have the expertise that goes with the normal operation of their cognitive faculties, broadly construed, and that gives their beliefs what authority they have for others, but those we honour with the title 'expert' or 'authority' have special skills or status that require a further explanation. The existence of such authorities is often connected with institutions rather than communities simpliciter. In both, epistemic authority can be exercised, formally in the one and less formally in the other, though there is an interaction between the two that can take positive or negative forms, involving either endorsement or opposition.

Zagzebski does not discuss institutional authority directly but it is something that brings a political dimension into the discussion of authority and autonomy in even the most relaxed, informal community contexts. It also, I think, provides more room for a degree of epistemic and practical caution to operate concerning the exercise of authority that goes beyond what Zagzebski usually allows. In the case of religion this means that there is room for tension as well as support between

what we might call a wisdom dimension and an institutional dimension to community authority. This relationship also calls into question any attempt to think of the authority of 'we' or 'us' in isolation from epistemic authority residing outside that community.

Zagzebski has an interesting discussion of wisdom and she points out correctly that it has, until very recently, received scant attention amongst contemporary philosophers. Recognition of wise people involves ceding a degree of epistemic authority to their views, but the institutional leaders of a religious community are not invariably wise people so that a tension may readily arise between their authority and that of the exemplars of wisdom in the community. The same is true, perhaps more acutely, of political authority. Short of wisdom, we can expect that communal authorities have some knowledge of areas relevant to their authority, indeed that is bound up with their *being* epistemic authorities, but where they occupy positions of institutional authority, it may become apparent that they lack not only wisdom but even the relevant knowledge. They also view themselves as guardians of the integrity of the institution and they are invested with the formal powers that go with institutional office. These facts expose them to the temptations of power and self-righteousness, and to the temptations of placing the need to 'save face' for the institution above the demands of justice (including epistemic justice – a topic discussed by Miranda Fricker) and accountability.[14]

A striking illustration of these dual temptations is provided within the Catholic Church by the alarming extent of clerical sex abuse of children as well as the 'Magdalene' incarceration and brutalising of young women in certain Irish convents (and elsewhere).[15] These tragedies were compounded by the appalling behaviour of the clerical authorities engaging in concealment, disingenuous denials, lies, and cover-ups when responding to information about those practices. Some of this involves the abuse of practical religious or political authority, but quite a lot of it involves the abuse or failure of epistemic authority, since the victims and the wider Catholic community trusted the clerical leaders to know what was right in doctrine and morals, to tell the truth about

[14] For Fricker's interesting views, see Miranda Fricker, *Epistemic Injustice: Power and Ethics of Knowing* (Oxford: Oxford University Press, 2007).

[15] The horrors of these convent practices have recently been dramatised in two films 'The Magdalene Sisters' and 'Philomena'. I cannot vouch for the detailed accuracy of these films, but there is little doubt about the factual situation from which their fiction (or 'faction') is derived.

how they acted on it, and to receive information in good faith. Where the abuses were not simply the result of outright vice and hypocrisy, as the individual sexual offences had been, the cover-ups and lies were often justified at the time by beliefs about the need to preserve the authority and reputation of the institution. In addition, the maltreatment of the women 'sinners' and their babies, for instance, were sometimes justified by authoritative religio-moral teachings about sex, about punishment for sin, and more broadly about extra-marital parenthood. Some of these beliefs were of course shared not only with many rank and file Catholics at the time but also with many in the wider community in Ireland and elsewhere. The involvement of parents and the Irish State in the Magdalene scandal supported the Church's role. Yet this itself really highlights some of the difficulties in isolating the religious authority of one community (whether epistemic or practical) from that of others.

The Catholic Church is not indeed unique in having such abuses perpetrated by its office-holders, and having higher authorities spurn the victims, protect the offenders, and guard the institution's reputation with secrecy, subterfuge and outright lies. On the list of abusive institutions investigated by the Australian Government's current Royal Commission on institutional sex abuse are: The Salvation Army, Scouts Australia, the YMCA, and an Anglican children's home. As this list shows, the problems are not restricted to religious institutions; besides the Scouts and the YMCA in Australia, in the UK the BBC and numerous public hospitals have been investigated by the police and other public bodies about their roles in the entertainer Jimmy Savile's ghastly depredations against young people over a period of 40 years.

A further feature of the clerical sex-abuse disaster is the degree to which the failures were abetted by reluctance to believe that members of the priestly fraternity were capable of such crimes. It is hard to know how much this factor worked with the authorities themselves, but, amongst other things, the reluctance indicates an excessive personal and epistemic trust (especially by parents) in the authority of religious leaders.

THE AUTHORITY OF 'EXTERNAL' COMMUNITIES

This brings us to another aspect of communal identity and the extended self which is that any given person is a member of many, often very different communities, external to the community in question and its sub-communities. And even when not directly a member of other

communities the knowledge held in those communities can be accessible to outsiders and relevant to their communal beliefs. So, in the case of religion, especially historically-oriented religions like Christianity and Islam, which are also religions of a book, the understanding of that religion's beliefs (core or otherwise) can be affected by knowledge or belief gained in scholarly communities such as those of history, philosophy, science and linguistics. Deference to the authority of such communities must have some impact for what survives the conscientious self-reflection of any member of the religious community.

In rejecting epistemic egoism, Zagzebski nonetheless insists that the autonomous self must be able to subject her reliance upon particular beliefs gained from authorities to 'critical self-reflection' (p. 228). Unlike epistemic egoism, this critical stance does not eschew all recourse to authoritative testimony nor accept only that which it can independently verify by its own individualistic resources, nor reject a role for emotion or the exemplary. But this means that it is not rigidly bound by its adherence to particular communal authority. I think Zagzebski does and should accept this much. But my argument goes further because I want to say that the communal authority itself is similarly open to such critical self-reflection and that opens the prospect of belief revision at the communal and institutional level as well. So the individual's rejection of some community belief or (in the institutional case) some official teaching may count as an effort to revise the community's self-conscious reflection and its results rather than, as Zagzebski usually puts it, a rejection of the community and its authority. Just when such revisions count as a rejection of what the community stands for or reveal a more authentic understanding of its reality is a difficult question, but it cannot be taken for granted that every such revision is a rejection. Consider the Irish child abuse scandals: Zagzebski's option of departure from the community is certainly one possibility, indeed it is the one which so many Irish Catholics have taken, but their departure is something the institution and many who remain in the Catholic community deeply regret. Not only that, but others, who are equally outraged by the beliefs and performance of the authorities, have stayed in the community vowing to reform the attitudes, beliefs and even structural features of the institutional authority, partly because of the role they regard these as playing in the scandals discussed briefly above.

A further illustration may be drawn from the history of Christianity. Christian unity was shattered in 1054, partly by a dispute about the

nature of the Trinitarian understanding of God. The Eastern churches split from the Roman churches over what is called the 'filioque' clause (in English, 'and the Son') in the version of the Creed called Nicene, though the phrase was not used at the Council of Nicea in 354 but adopted at later Councils. The amended Nicene Creed, after treating of Christ as sharing in the Divinity as the Son then proclaims: 'We believe in the Holy Spirit, the Lord, the giver of life who proceeds from the Father and the Son.' The Eastern churches rejected this since they held that the Holy Spirit and the Son both directly proceeded from the Father. This was a major doctrinal issue behind the Great Schism which created the Orthodox church traditions of the East as separate from Rome. Of course there were other issues, both theological and non-theological, as there always are. Many of these were political or quasi-political and I will say more of the role of the political later. But that differences over this wording should have had such momentous community consequences strikes me as amazing. I cannot understand why such an abstruse technicality as the difference over such 'proceedings' could be given a religious significance of this sort. I can understand how it might exercise a certain sort of theological and philosophical mind, but it seems to me irrelevant, hardly even reaching the peripheral, in terms of what matters in the Christian message. If the Nicene doctrine is a defining belief of the communal authority of the Church (not just the Catholic Church but most Protestant churches too) then critical reflection that rejects it or its importance surely calls for a change in what is of defining importance for membership rather than for change of community. It is, I think, significant that the other crucial Christian Creed in favour with most Christian denominations, including Catholic and some Orthodox, is the more ancient Apostles' Creed which makes no mention of internal relations within the Trinity.

I put this objection, probably not clearly enough, to Linda Zagzebski when she gave the lectures from which this book developed in Oxford in 2012. In the book, she addresses it in the context of a section on the need to resolve dissonance caused within a community by the fact that some other community which shares some of the first community's crucial beliefs also has a range of important conflicting beliefs that have for them survived conscientious self-reflection. She instances Christian and Moslem differences over the nature of God – Christians believing in the Trinity and Moslems rejecting that account in favour of a simple Unity. In response, she enunciates a *Need To Resolve Conflict Principle* as follows: 'It is a demand of rationality for a community to attempt

to resolve putative conflicts between its beliefs and the beliefs of other communities.'

She thinks that such conflicts may be resolved in various ways, such as changing some of our beliefs, or 'by adding a belief that explains why the dissenting community is mistaken, or by modifying the belief that conflicts with the belief of the other community in a way the members of the other community would not accept but which resolves our own dissonance' (p. 225). She also thinks that the fact that different communities have arrived at beliefs (perhaps different from our own) by conscientious reflection on the trustworthiness of those beliefs can in principle be recognised by each community and hence arises the possibility of advance in removing dissonance by inter-community dialogue.

This discussion of disagreement between communities is clearly relevant to some of my earlier comments and I will address it shortly. First, I want to see how it relates to the filioque issue and its role in the Great Schism. Zagzebski says that 'few theologians now consider it worth so much fuss' and concludes that the degree of the need to resolve dissonance within a community 'depends upon the degree of dissonance created within a community by the conflict. The degree to which a community cares about a belief is one dimension affecting degree of dissonance' (p. 225).

There are several things to say about this. The first is that the idea that 'few theologians now consider it worth much fuss', if true, as it may well be, needs to be set against the fact that the phrase and the doctrine it represents remain in the Nicene Creed which is recited daily in the Catholic Mass and other Christian denominational services. In so far as the amended Nicene formula, which include this account of 'the procession' of the persons of the Trinity, define what the Catholic and many other Western Christian communities believe then it seems they continue to care. In fact, the recitation may not have much if any cognitive and psychological resonance today, but then it is unclear how much the Western Christian communities cared about this in the 11th century. Many theologians cared, it seems, and apparently the Papal and Byzantine church and secular authorities cared, but we don't really know how the Catholic faithful felt. In any case, the question is whether all of them should have cared enough to split the Church.

The significance of the filioque clause is probably less doctrinal than political. The split of 1054 had been brewing since six centuries or more

as a conflict of power and authority between Rome and Constantinople. It embraced theological and liturgical issues as well, including the role of statues and icons, the use of unleavened bread in the Eucharist, but very prominently the primacy of the Pope in Rome. Serious historical research makes it plain that political motivations both within the religious institutions and in the wider political world played a major part in creating the widening gulf between the East and West even where they continued to have much in common doctrinally, and also impeded several genuine attempts at reconciliation.[16] It would be a mistake to discount sincere religious motivations, but also to discount the role of political motivations, since the two are often entwined.

If this is right, and I cannot marshal the evidence to support it fully here, then it illustrates the way that the institutional epistemic authority of a community is subject to scrutiny from within by those of its members who have reflected on a spectrum of considerations drawn from communal authority elsewhere which the conscientious self can (and perhaps must) attend to in its reflection on its trust in the institutional authority regarding community beliefs. So, understanding the history of the political and cultural forces at work in the controversy over 'filioque' can help the conscientious self determine whether this aspect of Trinitarian doctrine is or should be a crucial element of community belief. I have instanced here the role of historical authorities, but similar things could be said of biblical scholarship which has had an enormous impact on the way many Christians understand their faith today. Zagzebski gives one example of this about the accuracy of the Acts of the Apostles, but there are other cases more directly affecting doctrinal beliefs. Of course, there are various problems with the methodology of biblical criticism and there are sharp divisions within its ranks, a feature it shares with other disciplines in the Humanities. There are also, as Zagzebski notes, questions about whether some of the scholarly conclusions reflect prior metaphysical commitments that prejudge the evidence or whether the evidence independently supports those metaphysical views, for example,

[16] A good account of these political and religious factors can be found in Diarmaid MacCulloch, *A History of Christianity; the First Three Thousand Years* (London: Allen Lane, 2009). See especially Ch. 9, Part 1, and Ch. 10, Part 4. MacCulloch and other historians stress the political role played by the Emperor Charlemagne's conflict with Constantinople in bringing the matter to a head, even though at the time Pope Leo III refused to make the phrase part of the liturgy in Rome. This inclusion did not occur until the 11th century.

in the acceptability or rejection of miracles. But this merely emphasises the complexity of the materials that conscientious reflection must take into account.

POLITICS, CORRUPTION AND INSTITUTIONAL AUTHORITY

Another complicating factor in the case of institutional authority concerns the facts of corruption, especially the corrupting influences of power. Acton's famous dictum that 'all power tends to corrupt and absolute power corrupts absolutely' contains an insight into the temptations of institutional leadership and its authority that has been conspicuously vindicated by history. The more that the power of authority is centralised, not subject to close scrutiny, and remote from the desires, needs, and insights of those subject to it, the more its exercise can be distorted by aims inimical to the communal purposes which give it legitimacy. This is as true of the epistemological aspects as of the practical. The modern movement towards political democracy and its constitutional protection of citizen's rights against the abuse of power arises from an understanding that is indebted to the force of Acton's dictum.

In discussing Raz's account of political authority, Zagzebski comments that he has devised it as a contribution to the literature on political freedom within the framework of political liberalism, and she interprets this to mean that his account is constrained by a desire to maximise political freedom and to minimise political authority. She says that this perspective requires that 'it is more important to devise an account of authority that prevents tyranny than to give the bearer of authority the function of assisting the subjects in pursuing their individual and collective good' (p. 140). Consequently, she surmises that 'most modern political thought is motivated more by fear of bad authority than by desire for good authority' (p. 140). She gives no verdict on this liberal project, but insists that epistemic authority is different in that the oppression that the liberal political project aims to forestall rarely applies in the epistemic area, and hence Raz's liberal constraints are 'not important' for her project.

There may be some point in this dismissal for cases of non-institutional epistemic authority or even lightly institutionalised such authority, but once one realises that the exercise of most institutional authority involves considerable power vested in the authorities then the room for exercise

of that power in misguided, oppressive and corrupt ways is present, and the realisation of that fact raises a question both about the degree to which the authority has been distorted by these temptations and about the relevance of that answer to the trust to be given to the authority's claims, including knowledge claims.[17]

In the case of political authority, different people will no doubt answer these questions differently, but the widespread mistrust of politicians, even in liberal democratic societies, testifies to the verdict that most citizens have arrived at, and there is a considerable weight of history, not to mention the evidence of contemporary whistleblowers, on their side. This degree of mistrust is often lamented by commentators (and of course by politicians) and is said to impede good government, but it is unclear that more unqualified trust would remedy the situation.[18] Structural changes in the forms of authority and its implementation are more likely to lead to greater trust, though they may of course require trust in some authorities to devise them and to bring them about. Zagzebski's claim that most modern political thought is motivated by 'fear of bad authority than by desire for good authority' poses a false opposition since the desire for good authority should imply a fear of bad authority and the latter can motivate alertness to and respect for the former.

In the case of religious authority, it seems to me that points similar to the political 'constraints' apply as well. Indeed, the liberal and democratic spirit that informs the caution about political power and its relation to the individual's freedom should, it seems to me, also inform a similar caution about other forms of institutional power, including that in the area of religion whether the exercise of that power is practical, moral or epistemic. To return to the illustration of Catholicism, the present

[17] These distortions in the area of knowledge are often assisted by the control that political authorities exercise directly or indirectly over the language of communal discourse. Consider, for example, the way that the term 'enhanced interrogation techniques' has gained currency, especially in the United States, in the discussion of torture. Our enemies ('they') commit 'torture' where 'we' engage in 'enhanced interrogation techniques'.

[18] Political theorists have been much exercised by the nature of trust in political authority and institutions, the extent (if any) of its decline, and whether any such decline is a good or a bad thing. Notions like 'social capital' and 'civil society' are often bandied about in this connection, and often confusingly so. For good discussions of the complex problems in this area see: Mark E. Warren (ed.) *Democracy and Trust* (Cambridge: Cambridge University Press, 1999). Unfortunately, space precludes a recourse here to the insights of this debate.

Catholic structures of authority are still modelled on just those autocratic forms of political authority towards which the liberal project developed such well-justified suspicion. Those forms, I would argue, are ill-conceived in terms of a Divine mandate to maintain them at all costs, and it is interesting to see that Pope Francis is impatient with many of them, as so many ordinary Catholics have already become. But none of this is to say that the only resort for religious people is some form of epistemic and practical egoism. There remains a place for authority in religion, whether epistemic, moral and practical, and it will resemble the pattern of trust that Zagzebski so carefully and subtly develops. But it will have more stress upon the exemplary authority, it will require a more complex picture of community authority, and, while admitting the necessity of communal institutions, it will have what we might call both selective trust and selective mistrust in their operations and their office holders.

Zagzebski laments the 'disastrous effects' of the general decline of trust in authority in modern life and the denial of a role for authority as a condition for human fulfilment. She sees her book is an attempt to restore that trust in a modern form (p. 254). Yet such a modern form must admit that there are many areas of our lives where we rightly moderate the trust we give on the basis of knowledge we have gained from trusted others, for example, the cautionary trust many of us have in real estate agents and used car dealers, and yet this need not give ammunition to general epistemic or emotional egoism. So, a degree of selective mistrust for institutional authority, whether political, religious or moral, may be precisely what the executive self in its conscientious deliberations is rational to adopt. Whether the conscientious self adopts this stance will partly depend upon local facts available to that self, concerning, for example, the record of specific institutions, but is also likely to be influenced by broader political and social outlooks that themselves reflect both personal interests and experiences and also trust in sub-communal extended selves and exemplars. Consider the institution of policing. In liberal democratic societies people in comfortable circumstances and with dominant interests within the society that are mostly supported by the status quo tend to have an unqualified trust in the police institution, whereas the poor and disadvantaged tend to treat the representatives of that institution with more qualified trust. These sub-communities will tend to have quite different responses to complaints of serious police misbehaviour and to police denials of the allegations. The reactions

will be even more polarised where the sub-communities differ in racial or ethnic composition.[19] Of course, sound policing can be a crucial element in establishing and maintaining the rule of law and the security of citizens, but the opprobrium of the term 'police state' shows the potential downside of the institution and hence the need for caution about wholehearted institutional trust in this instance.

The idea of selective mistrust may however be thought to create a particular problem for a non-reductive analysis of our reliance upon the institution (as it were) of testimony. In particular, advocates of that analysis (or better, family of analyses) often point to the deep role of non-inferential trust in testimony that exists, necessarily, it seems in young children. Such children develop a framework of language and its concepts, and a basic set of reliable beliefs from parents and other adults early in life without which they could advance no further epistemically. It seems that this important stage proceeds with no recourse to anything like selective mistrust. But there are two comments to make on this. The first is that, as Thomas Reid insisted, this early state of tutelage is a preliminary to later stages that have a more sophisticated critical dimension, and it is then that selective mistrust plays a significant part. Reid says that our immature judgement 'is almost entirely in the power of those who are about us in the first period of life. If children were so framed as to pay no regard to testimony or authority, they must, in the literal sense, perish for lack of knowledge'. He goes on to say: 'But when our faculties ripen, we find reason to check that propensity to yield to testimony and to authority ... ', adding, that nonetheless 'the natural propensity still retains some force'.[20] Second, the picture of even very young children as totally passive epistemically in the face of adult testimony is, in any event, an unrealistic one, as most parents realise and much empirical work confirms. Quite young children ask for explanations, realise that some things they are told contradict others, and occasionally find that they have been misinformed where one informant testifies contrary to another. So, although their reliance on testimony goes very deep and helps indicate its significant role in adult life, thereby offering support

[19] The asymmetry of generalised trust in political institutions between groups with high socio-economic status and those with lower SES is argued for by the sociologist Orlando Patterson in his 'Liberty against the democratic state: on the historical and contemporary sources of American distrust', in Warren, op. cit., p. 196.

[20] Thomas Reid, *Essays on the Intellectual Powers of Man*, Derek Brookes (ed.) (Edinburgh Edinburgh University Press, 2002), Essay VI, Chapter V, section 10.

to the non-reductionist thesis, the example of early childhood learning also shows the incipient presence and emergence of selective mistrust.[21]

My remarks about religious and political institutions and the idea of selective mistrust suggest that the crucial concept of self-trust that Zagzebski employs so effectively may also need qualification, or, at least, further clarification. The complications of the extended self are less conspicuous for the individual self, but they are not entirely absent. Modern psychology has made much of divisions within the self, but they were also known to the ancients. Zagzebski realises there can be dissonance within the conscientious self and makes the need to resolve it a significant element in her project. But it is also characteristic of personal growth not only to seek the resolution of dissonance in our beliefs, but to discover the possibilities of mistrusting oneself: the healthy exercise of our cognitive and emotional powers should lead to the realisation that these capacities are not only essential to navigating our world and flourishing in it, but also prone to lead us astray in a variety of ways. Indeed we might say that the capacity for trust requires the robust capacity for (selective) mistrust as its other face.[22] This may lead to trusting others more than we trust ourselves, but it may also lead to seeing that some trusted authorities have been led astray by forces and defects we have recognised to mistrust in ourselves. These misled authorities may well be those of (one of) our own communities and its institutional leadership, and insight into this, and the tendencies contributing to it, may come from our trust in authoritative knowledge gleaned from communities quite outside our own, as well as those in other communities to which we belong.

[21] For samples of recent empirical work on the relevant capacities of very young children, see: 'Preschoolers' Search for Explanatory Information Within Adult-Child Conversation', Brandy N. Frazier, Susan A. Gelman, and Henry M. Wellman, Child Development, Vol. 80, No. 6 (2009), 1592-1611; and Paul L. Harris and Melissa A. Koenig, 'Trust in Testimony: How Children Learn about Science and Religion', Child Development, Vol. 77, No. 3 (2006), 505-524. My thanks to another (to me anonymous) contributor to this volume for raising the issue of child testimony in a comment on this chapter. My thanks also to Margaret Coady for alerting me to the scholarly literature on the critical capacities of young children.

[22] I am indebted for this thought to a comment by Karen Jones on an earlier draft of this paper.

AUTHORITY AND TRUST: REFLECTIONS ON LINDA ZAGZEBSKI'S *EPISTEMIC AUTHORITY*

JOHN COTTINGHAM

University of Reading

I. INTRODUCTION

Our modern egalitarian and individualistic age is suspicious of authority, and in recent times there have been almost daily reports in the press of cases where trust in various authorities, including financial, governmental, political and religious, has been found to have been abused or misplaced. Such disappointments seem to bolster the case for withholding trust in external authority and falling back on one's own resources. But if the lessons from Linda Zagzebski's groundbreaking work are accepted,[1] self-reliance turns out to be a confused and probably incoherent ideal (this is the critical or negative part of her thesis); and (more positively) the rational and self-reflective person is committed to believing and acting on authority. In the second half of this short discussion paper I shall raise some possible concerns about Zagzebski's positive case for reliance on authority, focussing on the moral and religious spheres. First however, let me say something about the negative part of Zagzebski's work, her critique of self-reliance. Since I find this wholly convincing, I shall confine myself to some supplementary observations, mainly to do with the historical context in which her critique is located.

II. FORWARDS AND BACKWARDS FROM KANT

One of the many virtues of *Epistemic Authority* is the light it casts on the genealogy of our modern philosophical culture. Zagzebski identifies Kant as the pivotal figure here; but her analysis goes far beyond the standard acknowledgment of his role in that upheaval in thought we know as the

[1] Linda Trinkaus Zagzebski, *Epistemic Authority: A Theory of Trust, Authority, and Autonomy in Belief* (Oxford: Oxford University Press, 2012).

'Enlightenment'. Kant's notion of the rational will as *selbstgesetzgebend*, 'giving the law to itself',[2] is often taken to be crucial in the modern shift from understanding morality as obedience to law to understanding it as self-governance.[3] But Zagzebski brings out what one might call the 'Janus-faced' character of this concept of the rational will: it looks backward to the traditional idea of reason as the foundation of authority, yet also forward to conceptions which accord primacy to individual choice as the ultimate source of normativity.[4] To put the matter more precisely, what gradually gathers speed after Kant is not just a shift from a rationalistic to a voluntarist conception of the basis of morality, but a further shift from understanding morality as grounded in something outside the self, which I am required to acknowledge as constraining my will, towards thinking of it as depending in the end on no more than my own individual decisions, or my own chosen self-conception.

The effects of this shift are clearly discernible, I think, in the language that has become so characteristic of the moral philosophizing of recent times. At first sight, to be sure, much of this language appears to accord primacy to the authority of reason, as for example in the 'constructivism' of Christine Korsgaard, which seeks to ground objectivity by reference to the rational procedures whereby we arrive at moral conclusions. But this approach still leaves open the question of what gives the moral values and maxims so arrived at their normative clout, or their ultimate authority over us. Korsgaard's answer to this, in the end, appears to be that if we were to violate them we would lose our sense of integrity and self-worth;[5] and here she seems partly to echo an earlier suggestion of Bernard Williams that the normative force of obligations derives ultimately from 'the *ethos,* the projects, the individual nature of the agent'.[6] Yet, as recent critics have

[2] Autonomy, for Kant, is 'the basis of the dignity of human nature and of every rational nature', according to which our will must be considered as *selbstgesetzgebend* ('giving the law to itself'). *Groundwork of the Metaphysic of Morals* [*Grundlegung zur Metaphysik der Sitten*, 1785], Ch 2; Akademie edition (Berlin: Reimer/De Gruyter, 1900–), Vol. IV, pp. 436, 431; transl. T. E. Hill Jr and A. Zweig (Oxford: Oxford University Press, 2003), pp. 236, 232.

[3] A shift traced out by J. B. Schneewind in *The Invention of Autonomy* (Cambridge: Cambridge University Press, 1998), cited in Zagzebski, *Epistemic Authority*, p. 19.

[4] Cf. Zagzebski, *Epistemic Authority,* pp. 23ff.

[5] Christine Korsgaard, *The Sources of Normativity* (Cambridge: Cambridge University Press, 1996), p. 102.

[6] Compare Bernard Williams, *Shame and Necessity* (Berkeley: University of California Press, 1993), Ch. 5, p. 103.

pointed out, putting it this way seems to get things backwards. The reason I couldn't with integrity live with myself if I betrayed a comrade is not that I have a certain conception of myself I can't give up; rather it is because I recognize something objectively morally repugnant about betrayal.[7] Or if this is denied, and my own self-conception or 'self-constitution'[8] is supposed to be the bedrock on which normativity rests, the question arises as to why some values should take precedence over others in determining how I conceive of or constitute myself. Though it is not her purpose to analyse these recent debates, Zagzebski's framework seems to me to cast a great deal of light on what is driving them, especially when she identifies a gradual 'degeneration' from the (initially arguably benign) Kantian conception of the rational will to a Nietzschean-style conception in which 'my own will, unconstrained by anything, including reason, is the only authority over me'.[9]

Moral self-reliance, construed in this latter way, turns out to be a very dubious notion, and one which, as Zagzebski shows, cannot properly be laid at Kant's door. For Kant himself insists (in a little noticed passage whose importance Zagzebski highlights) that even in the case of mathematical judgments, a position of 'logical egoism' is untenable: one's judgement must be tested by reference to the reason of others, if the individual is to have an 'external criterion of truth' (*criterium veritatis externum*).[10] This idea seems to me to be echoed in Kant's *Tugendlehre,* where parallel considerations apply in the moral sphere, to the deliverances of conscience. At first the picture may look rather subjective and individualistic, with Kant talking of conscience in terms of an 'internal forum' where my acts are brought before the tribunal of reason. But he goes on to argue that it is absurd to think that someone who is accused can be the same person as the one who judges; and hence the subject must think of himself as being judged by *another*, who is 'an ideal person that reason creates for itself'. Reason is thus abstracted from the contingencies of individual choice, and indeed comes close

[7] See Thomas Ritchie (following Thomas Nagel), in *From Morality to Metaphysics* (Oxford: Oxford University Press, 2012), pp. 96-7 and 101-2.

[8] See Christine Korsgaard, *Self-Constitution* (Oxford: Oxford University Press, 2009), Chs 1 and 2.

[9] Zagzebski, *Epistemic Authority*, p. 23.

[10] Immanuel Kant, *Anthropology from a Pragmatic Point of View* [*Anthropologie in pragmatischer Hinsicht,* 1798], transl. R. Louden (Cambridge: Cambridge University Press, 2006), p. 11; cited in Zagzebski, *Epistemic Authority*, p. 26.

to the religious idea of a supremely authoritative and rational judge. Notwithstanding this move towards an external criterion, however, Kant (consistently with his general rejection of speculative metaphysics) refuses to be drawn on the question of whether there is any really existing counterpart corresponding to the rational 'Other'. He considers the possibility that a human being is 'entitled, through the idea to which his conscience unavoidably guides him, to assume that such a supreme being actually exists'; but he pulls back from such a conclusion, and ends up saying that the idea of a supreme 'scrutinizer of all hearts' is given 'not objectively, but only subjectively'.[11] There is no space to evaluate the complexities of Kant's position here, except to observe in passing that his manoeuvre seems in the end to fudge the question of the authority of conscience. For once the actual existence of a supreme external authority is put to one side, it is unclear how the requisite authority can be furnished simply from my own resources. The ambiguity of Kant's legacy here is something moral philosophers are still wrestling with in one form or another.[12]

However that may be, I think it is worth noting that Kant's strictures against 'logical egoism' in some ways prefigure an insight later developed by Ludwig Wittgenstein, who perhaps surprisingly receives only one passing mention in *Epistemic Authority*. Wittgenstein surely deserves credit for definitively overturning the 'Cartesian' paradigm of the lonely epistemic inquirer who seeks the truth entirely from his or her own interior resources; for Wittgenstein's famous strictures against linguistic privacy entail that no thought or reflection would be possible in the first place unless it operated against the background of a public rule-governed network of language. This is effectively flagged up at the very start of his *Philosophical Investigations*, where Wittgenstein presents us with Augustine's (on reflection absurd) picture of the infant, fully equipped with reflective thought, who manages to work out what its parents mean when they speak by correlating their utterances and gestures with the various objects they refer to.[13] But the reality, of course, is that thought

[11] Immanuel Kant, *The Metaphysics of Morals: Doctrine of Virtue* [*Metaphysik der Sitten: Tugendlehre,* 1797], ed. M. Gregor (Cambridge: Cambridge University Press, 1996), Part II, §13.

[12] For an attempted defence of Kant on this issue, see Korsgaard, *Sources of Normativity.*

[13] Ludwig Wittgenstein, *Philosophical Investigations* [*Philosophische Untersuchungen*, 1953], transl. G. E. M. Anscombe (New York: Macmillan, 1958), §1, citing Augustine of Hippo, *Confessions* [*Confessiones c.* 398], I, 8.

and linguistic competence develop alongside each other, as the child is gradually inducted into a pre-existing culture. All of us grow towards understanding and awareness out of a position of complete dependency, by being bathed in a milieu of interactive intentionality that we cannot but trust and rely on, as inevitably and spontaneously as we rely on the food we eat or the air we breathe.

This basic climate of trust is by no means suspended even in Descartes' scenario of the lonely meditator setting out on the quest for knowledge; and here Zagzebski seems to me entirely correct in saying that the Cartesian method of doubt should *not* be construed as a justification of self-reliance.[14] It is of course true that the perspective adopted in Descartes' most famous work, the *Meditations*, is that of the solitary thinker, cut off from all contact with the outside world, and immersed in his or her own reflections. But the 'ideas' the meditator reflects on nevertheless have a publicly accessible structure; they are not dependent on the subjective psychological character of the meditator's experience, but relate to those 'immutable and eternal essences' which Descartes insists are quite independent of his own mind.[15] And the structure that grounds the objectivity of the essences so represented is none other than the mind of God – something as independent of the vagaries of any given individual's psychology as one might wish. Construing Descartes' epistemic stance in a wholly subjectivist way is only possible for the interpreter who implicitly secularizes Cartesian thought. If the stable, divinely underwritten structures of reason and meaning are set aside, and the meditator is left adrift in the isolated world of his own psychology, then it is hardly surprising that the whole Cartesian enterprise looks as if it is supposed to work in an entirely private domain. But that is not Descartes' way. His own philosophical journey is one which, in the very act of striving to break out of his self-imposed ordeal of doubt and uncertainty, comes up against an objective reality that is

[14] Zagzebski, *Epistemic Authority*, p. 17.

[15] René Descartes, *Meditations* [*Meditationes de prima philosophia,* 1641], Fifth Meditation, AT VII 64: CSM II 45. In this paper, 'AT' refers to the standard Franco-Latin edition of Descartes by C. Adam & P. Tannery, *Œuvres de Descartes* (12 vols, revised edn, Paris: Vrin/CNRS, 1964-76); 'CSM' refers to the English translation by J. Cottingham, R. Stoothoff and D. Murdoch, *The Philosophical Writings of Descartes*, vols I and II (Cambridge: Cambridge University Press, 1985), and 'CSMK' to vol. III, The Correspondence, by the same translators plus A. Kenny (Cambridge: Cambridge University Press, 1991).

the source not just of his own existence, but of those 'countless ideas' relating to the 'determinate essences, natures or forms', which are 'not invented by me or dependent on my mind'.[16] Even in the solitary mode of the First Meditation, the meditator's method of doubt could not even be formulated on the basis of the kind of private assignment of meanings which Wittgenstein famously attacks. 'Whether I am awake or asleep, two and three added together are five, and a square has no more than four sides.'[17] The objective constraints of language and meaning are in operation from the outset, and become ever more prominent as the argument of the *Meditations* develops.[18]

What thus emerges, with regard to Descartes, is that properly interpreted he is no advocate of epistemic self-reliance, since he philosophizes from the outset within a rational and objective space of meaning guaranteed by God. And if we tie this result in with the subsequent philosophical developments referred to above, as we trace the story from Kant and finally on to Wittgenstein, there seems to be a remarkable convergence: all three philosophers, properly interpreted, reject doxastic egoism or individualism, and insist that our belief formation must operate in a way that is subject to authoritative and objective constraints. This puts Zagzebski's critique of epistemic self-reliance squarely in line with the thinking of three of the most significant authors to have influenced the shape of modern philosophy.

III. AUTHORITY AND CONTINGENCY

Let me now very briefly outline Zagzebski's positive case for relying on authority, as I understand it. Her basic approach may I think be called a Cartesian one, if that often pejoratively employed term may be used for once without prejudice and in a purely methodological sense: her method is to 'proceed wholly from the point of view of the subject – a self-reflective person who asks herself how she should get beliefs she

[16] See citation in previous note.

[17] Descartes, First Meditation. The doubts subsequently raised by the introduction of the demon are, in my view, much weaker in scope than is often supposed. See John Cottingham, 'The Role of the Malignant Demon', *Studia Leibnitiana*, Vol. 8 (1976), pp. 257-64, reprinted in G. Moyal (ed.), *Descartes: Critical Assessments* (London: Routledge, 1991), Vol. II, pp. 129ff.

[18] For more on this, see John Cottingham, *Cartesian Reflections* (Oxford: Oxford University Press, 2009), esp. Chs. 5, 6, 13.

accepts upon reflection'.[19] This does not at all imply a commitment to full epistemic self-reliance, which turns out to be an untenable notion (for reasons discussed above); the point, rather, is that any philosophizing has to start from a position of self-reflective consciousness, where a certain degree of trust in oneself, and one's basic faculties, is rational and inescapable.[20] The next step involves a kind of presumptive epistemic egalitarianism: others clearly have the same basic faculties as I do, and given that I trust my own faculties, then prima facie I should trust theirs. So there is 'a general presumption in favour of the veridicality of the deliverances of the faculties of other persons until shown otherwise'.[21] Again, this egalitarianism has not a little in common with that of Descartes, who asserts in the opening sentence of the *Discourse on the Method* that 'good sense is the best distributed thing in the world'. He is no epistemic solipsist, but thinks of his method as available to anyone who is prepared to follow the path of rational inquiry, using the 'natural light' that is the normal birthright of every human being.

So far so good, though as suggested in the previous section I think we may usefully add to this basic premise about self-trust a further rider in the spirit of Wittgenstein. All philosophical inquiry must necessarily operate within an objective 'space of meaning' that has to be taken as given; no conscious self-reflection, even of a purely 'subjective' kind, could occur without presupposing a stable domain of logical and semantic rules. In trusting my faculties, as Zagzebski urges us to do, I have to entrust myself to this domain, in terms of which all my thinking and reflection must necessarily operate.

So far, again, so good. But now comes the problem for any theory which wishes to vindicate reliance on authority: what supports the domain itself, the fundamental set of conceptual structures to which I must entrust myself and which I must take as normative in my reflections? What gives this domain its objective authority over me, and what entitles me to assume that in employing my intellectual faculties and entrusting myself to it I will not be led astray?

Considered as a self-contained puzzle in epistemology, the problem may be intractable (one only has to think of the centuries of inconclusive wrangling over the 'Cartesian circle'). But if one looks at the question

[19] Zagzebski, *Epistemic Authority*, p. 2.
[20] Zagzebski, *Epistemic Authority*, p. 3.
[21] Zagzebski, *Epistemic Authority*, p. 185.

from a metaphysical standpoint, that is to say, if one puts the epistemic question aside and simply asks how our basic logical and semantic intuitions are pictured within the theistic and the secular worldviews respectively, then a great gulf opens up. The kind of traditional theism espoused by Descartes holds, as we saw in the previous section, that my intellectual faculties give me access to an eternal rational realm of 'essences and forms', held in place by the divine nature, in which my own nature has a dim and imperfect share. When I use my intellect conscientiously, focusing on the 'clear and distinct perceptions' disclosed by the God-given light of reason, I cannot be in error; for 'a reliable mind was God's gift to me'.[22] By contrast, if we move down to the secularist outlook that typically informs present-day philosophizing, the picture is very different. The framework that bestows objectivity is not an eternal framework stemming from divine reason, but is simply the framework of human culture, subject to all the contingency and change that is characteristic of our species and its history. Yet it is by no means clear that this is a worthy object of our trust, given what we know of the mistakes and confusions that have been handed down to us from our forebears, often encrusted with solemn assurances as to their irrevocable validity. Descartes' own warnings are instructive here: though he places his trust in the 'natural light' of reason, which leads him to assent to the clearest perceptions of the God-given intellect, he is anything but trustful of the *praejudicia*, the preconceived opinion or prejudices handed down from parents and teachers.

It is 'natural' for us to trust our most basic faculties. But 'nature', as Descartes explains in an important passage in the Sixth Meditation, is an ambiguous term. It can mean the authoritative order established by God, and the 'natural' light can mean the divinely instituted intellectual faculty that puts us in touch with that order. But in another sense it can simply mean the human impulse to jump to unwarranted conclusions, or to rely too readily on defective and suspect informants, be they the fluctuating and unreliable testimony of the senses, or the preconceptions we imbibed from our parents and teachers as children.[23] Trust is warranted only with respect to what is clearly and distinctly perceived

[22] Descartes, *Conversation with Burman* (1648), AT V 148: CSMK 334. The phrasing is as reported by Burman, but Descartes' own published formulations are closely similar; see Second Replies, AT VIII 144: CSM II 103.

[23] Descartes, Sixth Meditation, AT VII 80-83: CSM II 56-7.

(which is not to say Descartes thinks it is always easy to establish whether a given perception meets this standard).[24]

Now it is interesting to notice when thinking about the 'natural light' that, for Descartes, there is exact parity between the *ratio veri* and the *ratio boni* – the intuitions of the intellect with respect to logic and mathematical truth on the one hand, and with respect to moral goodness on the other. The light of reason discloses reasons of goodness that tell me that a certain value is to be pursued just as clearly and distinctly as it discloses reasons of truth that tell me that a mathematical proposition like 'two plus three is five' is to be affirmed.[25] How, then, might the moral secularist attempt to vindicate the authority of our basic intuitions in these two types of case?

A possible way forward with respect to our basic logical and mathematical faculties might be to point out that it is not possible coherently to mistrust them. This will not, of course, be a non-circular justification, but at least it brings us up against a bedrock to which there is no alternative; as Thomas Nagel has recently put it, in the case of our most basic logical reasonings, the only thing to think is that I have grasped the truth directly.[26] In the case of our moral perceptions, by contrast, there seems to be no question of a similar inevitability. And it is at this point that the historical and cultural contingency of our moral systems seems particularly worrying for any defender of an appeal to the legitimacy of authority in the moral sphere. Consider, for example, the recent attempt by John McDowell to establish a normative framework for ethics on the basis of what he calls 'second nature', that is to say, the complex nexus of moral sensibilities and propensities that have arisen through the development of human civilization and culture. For McDowell, these are perfectly 'natural', in the sense that they were developed out of our ordinary contingent activities as biological and social creatures of a certain kind, and hence they do not require us to posit any transcendent or supernatural properties or entities. But he argues that there are nonetheless genuine ethical reasons and requirements, to which we gain access by being inducted as children into a certain ethical culture; and in virtue of the access thereby gained, we do indeed, according to McDowell, become subject to moral requirements and demands. As he puts it:

[24] Descartes, Seventh Replies, AT VII 511: CSM II 348.

[25] Descartes, Fourth Meditation, AT VII 58: CSM II 40.

[26] Thomas Nagel, *Mind and Cosmos* (Oxford: Oxford University Press, 2012), p. 80.

> the rational demands of ethics are not alien to the contingencies of our life as human beings ... Ordinary upbringing can shape the actions and thoughts of human beings in a way that brings these demands into view.[27]

McDowell's position is a rich and subtle one which it would take us too far round to evaluate here. But the crucial point for present purposes is that, on the McDowell view, the 'reality' of the moral demands to which we are subject is in the end simply a function of a contingently evolved set of human characteristics and dispositions, and a contingently formed culture with a given developmental and social history. There is no further, no more ultimate, moral reality to constrain it or measure it against, and no ultimate *telos*, no objective goal that represents the final purpose of human ethical life. The developmental history of our species and the genealogy of our ethical culture is a contingent one; it might have been otherwise, and if it had, then, it seems to follow that the relevant ethical 'realities' and 'demands' might have been different. The potentially subversive implications of this kind of picture were acutely discerned by Bernard Williams when he spoke of the 'radical contingency of the ethical'.[28] The problem, in a nutshell, is that once the idea is accepted that the authority and power of the moral demands which seem to call forth our allegiance depends on our past history and the culture into which we happen to have been inducted, then true normativity evaporates. The 'morality system' becomes one among other potential systems, a 'peculiar institution',[29] whose shackles we may think (as Nietzsche for example did)[30] that we have reason to shake off in our quest for self-realisation or some other alluring project.

It may seem strange to invoke these secularist pictures of ethics when discussing the views of Zagzebski, who as a theist will be committed,

---

[27] John McDowell, *Mind and World* (Cambridge, MA: Harvard University Press, 1996), p. 83.

[28] '[A] truthful historical account is likely to reveal a radical contingency in our current ethical conceptions. Not only might they have been different from what they are, but also the historical changes that brought them about are not obviously related to them a way that vindicates them against possible rivals.' Bernard Williams, *Truth and Truthfulness* (Princeton: Princeton University Press, 2002), Ch. 2, p. 20.

[29] See Bernard Williams, *Ethics and the Limits of Philosophy* (London: Collins, 1985), Ch. 10.

[30] See Friedrich Nietzsche, *Beyond Good and Evil* [*Jenseits von Gut und Böse*, *1886*], §37 and §203. See also John Cottingham, 'The Good Life and the "Radical Contingency of the Ethical"', in D. Callcut (ed.), *Reading Bernard Williams* (London: Routledge, 2008), Ch. 2, pp. 25-43.

I take it, to the existence of an eternal, non-contingent, domain of moral reality, and who would therefore repudiate both the Williams-style conception of the 'morality system' as a purely contingent product, and the McDowell-style project of grounding moral normativity in the historical and social facts of human nature and culture. But what these alternative pictures show, I think, is that Zagzebski's argument will be incomplete unless it explicitly moves beyond epistemology to metaphysics, and offers an account of the nature of moral truth, and of our reflective moral beliefs as apt to track that truth. For if the argument for relying on moral authority is conducted in purely epistemic terms (to do with conscientious belief formation), the worry will remain that in enlisting the expertise or even the advice of others, as Zagzebski urges, I may simply be appealing to the propensities and sensibilities of those whose faculties, like mine, have been shaped and formed by induction into a certain culture with contingent historical roots, which might have been otherwise, and which might well come to be superseded. It does not seem enough here to reply that moral authority is justified for the subject by her conscientiously judging that she is more likely to get a belief that will satisfy her future conscientious reflection if she takes the belief on the authority of the others than if she does not.[31] For the deliverances of further conscientious reflection are surely themselves likely to be shaped by the very system whose authority is here in question. And if this is right, the case for moral authority seems to be threatened. One thinks here of Sabina Lovibond's discussion of how far the self-conscious aspirant to virtue is in a position to respond to challenges about the merits of the currently prevailing process of ethical formation. The worry is that 'a map of the domain of value ... cannot be drawn with any authority in advance of finding answers to the very questions with which it is meant to help us'.[32]

A parallel kind of argument seems to me to apply, *mutatis mutandis*, to the case of religious authority, where there seems to be an obvious analogue of Bernard Williams' idea of the 'radical contingency of the ethical'. So a devout Roman Catholic from Dublin or Boston, for example, may reflect that had he been born in Bagdad or Tel Aviv, and inducted into an Islamic or a Jewish culture respectively, he might have been inclined

[31] See Zagzebski, *Epistemic Authority*, p. 199.

[32] Sabina Lovibond, *Ethical Formation* (Cambridge Mass: Harvard University Press, 2002), p. 188.

conscientiously to judge that his religious beliefs would be more likely to satisfy his future conscientious reflection by his accepting the authority of the Islamic or Jewish tradition – and all this with the same degree of sincerity and conscientious commitment that characterises his present deference to Catholic authority. In an insightful and humane chapter on 'Trust and Disagreement', Zagzebski does not shrink from confronting the tensions that can arise when I disagree with someone I respect from another faith (for example about the trinitarian nature of God), while at the same time recognizing that I have a prima facie reason to trust their conscientious self-reflection as much as my own. Her answer, if I understand it correctly, is that there is no easy way to resolve 'the antinomy of reasonable disagreement', but that I simply have to balance my estimations of the relative trustworthiness of those of my own faith community, and the likelihood that on reflection I will continue to be able to identify with them, against the possibility that by making radical changes to my emotional and doxastic outlook I can develop a new religious allegiance that will better survive future conscientious self-reflection. Such a change of allegiance might in certain circumstances be the appropriate course, but given the adjustments that would be needed, in most cases the reasonable and conscientious option will be to stay where I am.[33] Although, in fairness, this is not offered as a full resolution of the antinomy of reasonable disagreement,[34] I cannot but feel some reservations over the proposed framework of conscientious self-reflection as a means of approaching it. For, as in the moral case, once I fully and deeply acknowledge the contingencies of my present allegiances, and the extent to which my honest self-reflection might have been different had I been inducted into a different culture, then the supposedly objective and rational epistemic basis for my deference to a given religious authority starts to look less secure.

These concerns, such as they are, do not in any way detract from my admiration for Zagzebski's project, or for the superb philosophical finesse with which she executes it. Her powerful exposé of the incoherence of epistemic egoism seems to me unassailable; and as one who shares her particular religious allegiance, I am strongly sympathetic in principle to her aim of vindicating moral and religious authority. She speaks at one point of 'intellectual humility' in connection with trusting a religious

[33] Zagzebski, *Epistemic Authority,* p. 221.

[34] Zagzebski, *Epistemic Authority,* Ch. 10, §4, opening paragraph.

community,[35] and this seems to me a highly persuasive part of her argument. It would indeed be a strangely perverse kind of arrogance to insist that if I try to figure out what to do all by myself, the outcome is more likely to survive my own future conscientious self-reflection than if I defer to the authority of a community that has flourished for many hundreds of years.[36] But the problem remains that in deferring to the authority of another I may be allowing myself to be overborne by the power of entrenched tradition or preconceived opinion, or, more radically, that the very sensibilities and responses that incline me to such deference may themselves have been shaped by induction into the tradition whose authority is in question.

The best way to break out of this impasse seems to me to proceed not by further epistemic theorizing but by action; that is, to follow the 'Pascalian'[37] suggestion of ceasing to be preoccupied from an external standpoint with beliefs and their justification, and instead moving inside a faith tradition – entrusting oneself to a structure of communal praxis that fosters moral and spiritual enrichment, opening the way to new and deeper kinds of experience which, if all goes well, I may come to see retrospectively as validating my act of trust. Here I would wholeheartedly agree with the more pragmatic or practically oriented note sounded by Zagzebski towards the end of her remarkable book, in the chapter on religious authority:

> There are other natural desires [apart from the desire for truth] that can be better satisfied by participation in a wisdom community than on one's own. These desires include the desire to know and to do the good, to acquire not just knowledge, but understanding, to learn patterns of living and principles of action that result in a more integrated self, to be surrounded by grace and beauty, and to experience the delights of living among persons whose own pursuit of these ends enhances one's own.[38]

This beautifully expressed set of aspirations exemplifies the kind of goal that cannot plausibly be achieved by a 'do-it-yourself' approach. Trust, to come back full circle to the point with which we started, is always open to abuse, and entrusting oneself to a tradition, or committing oneself to an individual or to a community, is never without risk. But without

[35] Zagzebski, *Epistemic Authority,* p. 199.
[36] Cf. Zagzebski, *Epistemic Authority,* p. 148.
[37] Cf. Blaise Pascal, *Pensées* [1670], ed. L. Lafuma (Paris: Seuil, 1962), no 418.
[38] Zagzebski, *Epistemic Authority,* p. 201.

such trust, we may be cutting ourselves off from the very possibility of arriving at the evidence needed to justify that trust, and of achieving the spiritual goods that vindicate it. If that is a paradox, it is a paradox which flows from the inescapable dependency that is part of what it is to be human.[39]

[39] I am most grateful to Fiona Ellis for helpful discussion of an earlier draft of this paper.

ZAGZEBSKI ON RATIONALITY

DUNCAN PRITCHARD & SHANE RYAN

University of Edinburgh Soochow University, Taipei

INTRODUCTION[1]

This paper examines Linda Zagzebski's (2012) account of rationality, as set out in her rich, wide-ranging, and important book, *Epistemic Authority: A Theory of Trust, Authority, and Autonomy in Belief*. We briefly describe the account that she offers and then consider its plausibility. In particular, in the first section we argue that a number of Zagzebski's claims with regard to rationality require more support than she offers for them. Moreover, in the second section, we contend that far from offering Zagzebski a quick way of dealing with radical scepticism, her account of rationality actually seems to be particularly vulnerable to this problem.

I. ZAGZEBSKI ON RATIONALITY

For Zagzebski (2012: 30), 'rationality is a property we have when we do what we do naturally, only we do a better job of it', while being rational 'is to do a better job of what we do in any case, what our faculties do naturally'. Zagzebski is considering rationality in a broad sense of the term, a sense related in a certain way to the relationship between aspects of the self, for example, beliefs and desires. In arguing that a standard of rationality is provided by how a self naturally operates, Zagzebski (2012: 33) writes that 'there is a connection between the natural and the normative, in particular, a connection between the self as it naturally operates and the way it should operate'. Zagzebski thus holds that the kind of performance that allows us to be rational is guided normatively by how we operate naturally.[2]

[1] We are grateful to an anonymous reviewer from the journal for detailed comments on an earlier version of this paper. Thanks also to Marcin Iwanicki and Linda Zagzebski.

[2] Note that Zagzebski (e.g., 2012: 30-31) sometimes uses 'automatically' and 'unconsciously' as synonyms for 'naturally'.

Zagzebski (2012: 29) notes that '[b]eliefs, emotions, desires, and decisions can conflict with one another'. When there are such conflicts, we may experience dissonance. In fact, the experience of dissonance is a way we have of detecting such conflict. Although for Zagzebski (2012: 31) in some cases we can get on fine without resolving dissonance, it's better if dissonance is resolved. Indeed, for Zagzebski a harmonious self is something we desire and attempt to achieve by our nature. Relatedly, Zagzebski argues that we also have a natural desire for the truth.[3] Conflicts can sometimes be resolved unconsciously or naturally. For example, you may believe that you have turned off your alarm clock, but then you hear the alarm going off. You unconsciously give up your belief that you have turned off the alarm clock. Sometimes, however, dissonance caused by conflict is not resolved in this way.

Aside from writing that being rational is doing a better job of what we do naturally, Zagzebski (2012: 217) specifically holds that resolving detected conflict in the self that one is aware of is part of being rational. This may involve, for example, the targeting of a belief to be given up in the face of a conflict between beliefs. She writes that when we engage in such targeting, we should ask ourselves which belief is most likely to withstand future self-reflection. So to be rational is to do a better job than what we do automatically, it requires resolving conflict in a way that we judge will survive future self-reflection, and it is better to resolve conflict given our natural desire for a harmonious self. It follows that an agent who, for example, resolves a conflict in their beliefs in a non-automatic way, and does so in a way which she judges will survive future reflection, is being rational.[4] For Zagzebski (2012: 231), although what we judge will survive future reflection might turn out to be wrong, there is a connection between our judgements of what will survive future reflection and the truth of what will survive future reflection.

Rationality is good because having such rationality means being better placed to have the harmony of self that we naturally desire and

[3] Zagzebski sees this particular natural desire as part of the pre-reflective self. Interestingly, it is common these days for epistemologists to claim that truth is not even a fundamental epistemic good (i.e., something which is good along a specifically epistemic axis of evaluation), let alone a value that we naturally desire. See Pritchard (2014b) for a recent discussion of some of the issues in this regard.

[4] Note that any such resolution of conflict would have to come about in an appropriate way; e.g., it would have to have to have the right aetiology.

naturally seek to achieve. There is a sense in which Zagzebski's claims about our nature being such, and about nature being a marker for the normative, leads to our conscious selves being required to be a certain way normatively.[5] Not only are we not being rational if we don't respond to detected conflict in the way that Zagzebski outlines, but we're also failing in a normative sense.

How plausible is Zagzebski's view of rationality? Why should we think that the natural is a guide to the normative? Zagzebski sets these positions out rather than providing a full account of rationality. Perhaps one thought here might be that we are naturally well-functioning, so when we come to consider how we should be in various domains, how we are pre-reflectively or unconsciously provides some initial guidance.[6] And while there may be some cases in which how we are naturally or what we do unconsciously should not be taken as providing normative guidance, for example our natural thinking with regard to probabilities, we might think such cases are the exception rather than the rule.

There are a number of empirical issues raised by Zagzebski's claims, perhaps the most fundamental of which includes whether humans naturally resolve or try to resolve conflicts between different parts of the self. One might wonder whether this is what we naturally do or whether something else, such as suppression of conflict in the self, is usual.[7] In fact, it seems plausible that frequently we don't naturally do anything about conflicts in the self; an agent gains a new belief that conflicts with an old one, or a belief that conflicts with a desire, and the self doesn't naturally do anything about the conflict. If it turns out that we don't typically naturally resolve conflicts in the self, or even naturally attempt to solve such conflicts, and we accept Zagzebski's (2012: 32-33) claim that how the self operates naturally with regard to conflict in the self serves as a standard for rationality, then why should we think that it is rational to resolve conflicts in the self? It would be good, therefore, to

[5] Interestingly, the requirement that we resolve any conflict by judging what will best withstand future self-reflection is a way of promoting diachronic self-harmony.

[6] Alternative thoughts might be that Zagzebski's positions enjoy intuitive plausibility, or that, given that they together with what she has to say about epistemic authority provide us with interesting and plausible answers on the topic of epistemic authority, it makes sense theoretically to adopt these positions.

[7] Zagzebski (2012: 30) mentions that the self may sometimes attempt to resolve conflict by suppressing the feeling of dissonance, which is a somewhat different claim.

have empirical evidence that supports the characterisation that naturally we resolve conflicts in the self.[8]

Even if it turns out that natural resolution of conflict in the self, or its attempt, is typical, it's not clear on what basis Zagzebski can appeal to only those cases and not other cases in which we operate naturally with regard to conflict in the self. She writes of there being a connection between the natural and the normative, but we need a basis for distinguishing between different natural operations, as presumably Zagzebski doesn't want to offer all natural operations with regard to conflict in the self as a standard for the rational, but only some. Perhaps she might support favouring the resolving of conflicts in the self over other natural operations by appeal to what she claims is our natural desire for harmony in the self.

Whether we have natural desires again is an empirical claim. Given that, as Zagzebski (2012: 33-34) acknowledges, philosophers in recent times generally don't claim that we have such desires, the dialectical burden is on her to give us some reason to think that there are such desires.[9] Furthermore, even if we do have basic desires, such as a non-reflective desire for truth and a desire for harmony in the self, it's not obvious that their origin is in the natural. Automatic or unconscious behaviour needn't have its source in nature. Some of what Zagzebski attributes to the natural may be non-natural, say, cultural in origin. It may even be the case that the societies that survive are ones that, say, place a high value on truth, and so there is universal or near universal acceptance of the value of truth.[10] It would be good to have a more detailed account of natural desires, including empirical evidence that we do desire what Zagzebski claims we desire and that those desires are natural desires, especially given the role the natural as a standard for the normative has on Zagzebski's account.

A possible nearby alternative to Zagzebski's view that we have the natural desires that she describes is that we are naturally desirous of

[8] Longworth (2013: 159) also notes the lack of empirical evidence for claims made by Zagzebski about the natural.

[9] Longworth (2014: 161) makes a similar point, writing that it would be good if Zagzebski offered some reason to think that contemporary philosophers do or should accept that there are natural desires. Zagzebski (2012: 33-34) does write that traditionally there has been discussion of natural desires in philosophy and that reference to natural desires is common in other disciplines, though it's fair to say that this falls short of giving us a reason to think that there are such desires.

[10] Or it may simply be the case that some cultures don't place a high value on truth and that there is not a universal or near universal desire for truth.

evolutionary success (i.e., we desire to survive and prosper, for our kin to survive and prosper, to successfully reproduce, and so on). A desire for evolutionary success may propagate derivative desires the fulfilment of which are instrumental to evolutionary success. Such examples may include being able to get food, form social bonds, navigate social hierarchies, and so forth. A desire for such successes may often happen to lead us to getting truths and behaving in consistent ways. The view presented here could explain much of the initial intuitive pull of the thought that we have a natural desire for truth and a natural desire for harmony. Given that this possible nearby alternative seems relatively obvious and at least as plausible as the natural desires that Zagzebski describes, it's an alternative that defenders of the view that we have the natural desires that Zagzebski describes should rule out.[11]

Even if what Zagzebski attributes to the natural is indeed natural with regard to us resolving conflict within the self and having the desires she describes, we have good reason to question that the natural can be a good guide to the normative. Presumably the way we naturally form beliefs, naturally respond to conflicts within the self, and the natural desires we have, are best explained as part of an evolutionary story. As such it would be odd to think that, for example, the way we've happened to evolve to respond to conflicts in the self is connected to how we should respond to such conflicts. After all, what we naturally do in these respects presumably is a product of what has been evolutionarily advantageous, but what is evolutionarily advantageous doesn't obviously coincide with what we have thought of as, say, epistemically normative. Furthermore, any story where the natural provides a guiding role to the normative in ways that go beyond saying roughly that we should promote the survival of our genes looks committed in principle to relativism. After all, natures other than our own are theoretically possible, there may be alien species that even have such natures. If nature is a standard for the normative, and if there can be several natures, then there can presumably be several normative standards, for example in the ethical domain or the epistemic domain. While this isn't obviously wrong, it's a consequence of Zagzebski's stance that warrants consideration.

[11] Alternatively, such defenders may claim that though there may be much overlap between the two sets of desires, we naturally have both sets of natural desires.

II. ZAGZEBSKI ON RATIONALITY: RADICAL SCEPTICISM

Zagzebski appeals to her view of rationality in order to offer a quick dismissal of the problem of radical scepticism. Roughly, her claim is that since rationality is doing a better job of what we do naturally, it follows that radical scepticism cannot be rational since it is not natural. (Zagzebski 2012: 45) While undoubtedly appealing, we do not believe that this approach to radical scepticism stands up to closer scrutiny.

The first point to note about this way of dealing with radical scepticism is that it confuses the *problem* of radical scepticism with the radical sceptical *position*, where the latter is a philosophical stance which embraces the radical sceptical conclusion. As many commentators in the contemporary debate about radical scepticism have noted – most notably Barry Stroud (1984) – it is very important to keep these two ways of thinking about radical scepticism apart. Part of the reason for this relates to the very point that Zagzebski makes, which is that we know in advance that the sceptical position is untenable in various ways – for example, we know that it would lead to cognitive paralysis and that it is almost certainly psychologically impossible.[12]

Merely noting the implausibility of the radical sceptical stance doesn't in itself offer us any intellectual comfort when it comes to resolving the radical sceptical *problem*, however, since the challenge posed by this problem doesn't in any way trade upon whether being an actual radical sceptic is a viable theoretical option. Indeed, the sceptical problem is best thought of as a putative *paradox* – *viz.*, a series of claims which are highly intuitive when taken individually, and which on the face of it are rooted in our ordinary epistemological commitments, but which collectively generate a contradiction. Radical scepticism *qua* position is then one particularly dramatic way of resolving this paradox, but the point is that any resolution will involve denying something intuitive (or else demonstrating that it is not as intuitive as we initially supposed).[13]

[12] Note that, like Zagzebski (though see note 13), we are specifically talking about *radical* scepticism here. In particular, there may be more moderate forms of scepticism (some varieties of Pyrrhonian scepticism, for example) about which this doesn't apply. For the claim that radical scepticism leads to cognitive paralysis, see, e.g., Wright (2004). For the claim that radical sceptical doubt is psychologically impossible, see Strawson (1985).

[13] Part of the reason why Zagzebski fails to notice this point could be her tendency to treat Pyrrhonian scepticism as representative for radical scepticism more generally, even while discussing features of the radical sceptical problem (such as the problem of epistemic circularity) which aren't essentially allied to the Pyrrhonian stance. Pyrrhonian scepticism – even in its most radical guise (see note 12) – is unusual, however, in that it

Note also that, as Stroud (1984) famously argued, radical scepticism *qua* paradox can be rooted in our everyday – and thus 'natural', by Zagzebski's lights – epistemological commitments and yet nonetheless consist of a kind of epistemic evaluation which is very different from our ordinary practices. The point is that the sceptical system of epistemic evaluation is meant to be a *purified* version of our everyday practices, one that abstracts away from irrelevant limitations of time, imagination, thoroughness, and so on. In this way, it can *both* be true that radical scepticism is rooted in our everyday epistemological commitments and that, for example – see Zagzebski (2012: 36) – we naturally trust our faculties to lead us to the truth.[14]

Imagine, if you will, that we follow our natural desire for the truth that Zagzebski outlines, but do so unfettered by purely practical concerns. The problem posed by radical scepticism is that we seem thereby to be led to a position according to which knowledge is almost impossible to obtain. Radical scepticism thus falls out of a perfectly natural way of reasoning. In particular, we are led to radical scepticism by doing what we naturally do on this score, albeit in a more thorough and exacting way. Since, for Zagzebski, the hallmark of rationality is to do what we naturally do, albeit in a better way, then it seems that the radical sceptical paradox is not the result of irrational ways of thinking, but rather a by-product of a view of rationality which, as in Zagzebski's account, is rooted in the natural.

The upshot is that Zagzebski's account of rationality does not offer the quick resolution of the problem of radical scepticism that she supposes. In fact, her account of rationality seems particularly vulnerable to this very problem.[15]

---

is an explicitly *embodied* form of radical doubt. It thus tends to obscure the distinction between radical scepticism *qua* position and *qua* paradox.

[14] Indeed, it is no part of radical scepticism on this construal to dispute that it is natural to trust our faculties. The point is rather that this trust is incompatible with some of our other natural epistemological commitments, such that something is amiss *somewhere* – that's more than enough to motivate radical scepticism, it does not need to be further added *where* our intellectual mistake lies. For further discussion of the very idea of radical scepticism *qua* paradox, see Pritchard (2014a; forthcoming, part one).

[15] Zagzebski (2012: 30-31) notes that some paradoxes – she mentions the Lottery Paradox and the Preface Paradox in this regard – are such that they don't provoke an experience of dissonance. Instead, we treat them as 'intellectual puzzles'. Perhaps radical scepticism *qua* paradox is of this kind (Hume famously thought so). But note that this doesn't mean that such paradoxes are intellectually idle. As epistemologists, we need to understand what is generating the paradox and thereby determine the means to resolve it, even if the paradox has no sway over our day-to-day intellectual lives.

BIBLIOGRAPHY

Baril, A. 2013. 'Review of *Epistemic Authority: A Theory of Trust, Authority, and Autonomy in Belief*, *Notre Dame Philosophical Reviews*, available at: <http://ndpr.nd.edu/news/41479-epistemic-authority-a-theory-of-trust-authority-and-autonomy-in-belief/> [accessed 28/11/14]

Longworth, G. 2014. 'Epistemic Authority', *Analysis*, 74: 157-166

Pritchard, D. H. 2014a. 'Sceptical Intuitions', in *Intuitions*, D. Rowbottom & T. Booth (eds.) (Oxford: Oxford University Press), pp. 213-31

Pritchard, D. H. 2014b. 'Truth as the Fundamental Epistemic Good', in *The Ethics of Belief: Individual and Social*, J. Matheson & R. Vitz (eds.) (Oxford: Oxford University Press), pp. 112-29

Pritchard, D. H. Forthcoming. *Epistemic Angst: Radical Skepticism and the Groundlessness of Our Believing* (Princeton, NJ: Princeton University Press)

Strawson, P. F. 1985. *Skepticism and Naturalism: Some Varieties* (New York: Columbia University Press)

Stroud, B. 1984. *The Significance of Philosophical Scepticism* (Oxford: Clarendon Press)

Wright, C. J. G. 2004. 'Hinge Propositions and the Serenity Prayer', in *Knowledge and Belief*, W. Loffler & P. Weingartner (eds.) (Vienna: Holder-Pickler-Tempsky), pp. 287-306

Zagzebski, L. T. 2012. *Epistemic Authority: A Theory of Trust, Authority, and Autonomy in Belief* (Oxford: Oxford University Press)

ZAGZEBSKI, AUTHORITY, AND FAITH

TRENT DOUGHERTY

Baylor University

INTRODUCTION

Epistemic Authority is a mature work of a leading epistemologist and philosopher of religion (and metaphysician, too, but that character doesn't feature in this story). It is a work primarily in epistemology with applications to religious epistemology. There are obvious applications of the notion of epistemic authority to philosophy of religion. For, on the face of it, the notion of some kind of 'epistemic authority' may serve as a conceptual anchor for our understanding of faith. Indeed, there is ample historical precedent for this. Faith, says Locke, is 'the assent to any proposition ... upon the credit of the proposer, as coming from God, in some extraordinary way of communication'.[1] In later Lockeans, 'credit' is often rendered 'authority', and the terms were used synonymously at the time of his writing.[2] One of the beauties of Locke's view is its reductionism, that is, it's parsimony, which is a species of elegance and therefore beauty. Zagzebski's notion is more high-octane than Locke's. In this essay I will do four things. In Section 1 I will describe two kinds or notions of authority or at least two usages of the word 'authority'. In Section 2 I will describe Zagzebski's use of one of these notions, the non-Lockean one, to ground the reasonableness of religious belief. In Section 3 I will give four arguments against her view. In section 4 I will reply to her critique of Locke. The upshot, in my view, is that though we learn much (very much indeed) from *Epistemic Authority* (about both testimony in general and religious testimony in particular among many other things), a more Lockean approach to the nature of faith is still preferable.

[1] Locke, *Essay*, Book IV, Chapter XVII, section 2.

[2] See, for but two examples, Leland (1740: 15), and Meadly (1809: 16).

EUROPEAN JOURNAL FOR PHILOSOPHY OF RELIGION 6/4 (WINTER 2014), PP. 47-59

I. TWO KINDS OF AUTHORITY

1.1 Expert Authority

The first kind of authority I shall call *expert* authority. This usage is usually met in a sentence such as 'She is an authority on horses', or 'He is an authority on the game of golf', or 'She is an authority on 19th Century Russian Novels', or 'He is an authority on first century Palestine'. In each of these sentences, the word 'authority' could be replaced with 'expert' without addition or loss. What we mean when we say such things is (at least) that so and so knows lots about that subject. And we also intend to convey (most of the time)[3] that that person is a reliable source of information – they are very likely to state the truth of the matter concerning that area – and so their testimony can be relied upon[4] (= we can trust them in the matter = believing them is warranted = their says-so is evidence that it is so).

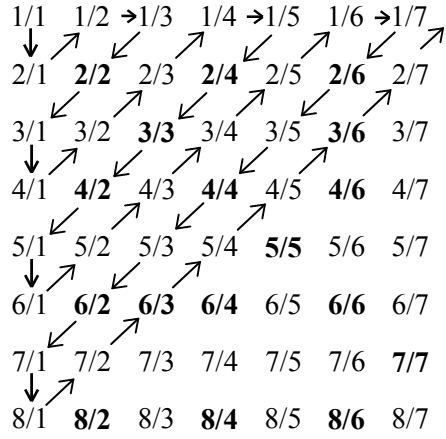
Notice, though, that if X is an expert it is still an open question whether and in what manner X's opinion, expert though it be, should affect your opinion. For you yourself may be an expert! And of course being an expert on some subject – taking the sentences above as paradigmatic – comes in degrees. So imagine two dials side by side. One dial represents A's expertise on a subject, another represents B's. Let the dials have ten 'clicks' labelled from '1' to '10'. Here are the possible pairs of settings representing the relationship between A's expertise to B's expertise.

{<1,1>,<1,2>,<1,3>,<1,4>,<1,5>,<1,6>,<1,7>,<1,8>,<1,9>,<1,10>,
<2,1>,<2,2>,<2,3>,<2,4>,<2,5>,<2,6>,<2,7>,<2,8>,<2,9>,<2,10>,
<3,1>,<3,2>,<3,3>,<3,4>,<3,5>,<3,6>,<3,7>,<3,8>,<3,9>,<3,10>,
<4,1>,<4,2>,<4,3>,<4,4>,<4,5>,<4,6>,<4,7>,<4,8>,<4,9>,<4,10>,
...
<10,1>,<10,2>,<10,3>,<10,4>,<10,5>,<10,6>,<10,7>,<10,8>,<10,9>,<10,10>}

[3] There are certain areas – like nutrition – where the people with the most knowledge tend to change their minds frequently or cannot make up their collective mind at any one time: 'Don't eat eggs! Eggs will kill you!' 'No, eat all the eggs! Eggs will save you!' Between this and matters of great expert agreement, there is a spectrum.

[4] Some people make much ado about the notion of reliance, but I do not. What I mean by reliance here has nothing to do with some kind of right to resent those who get it wrong. Rather, I mean something like if you used that person as a source of information to place bets in the relevant field, you'd come out ahead in the long run.

This set of ordered pairs describes a 'discrete continuum' (if you will) between two individuals with no expertise (if we let 1 represent the average base unit of knowledge) to two individuals who are each experts. Or we could think of the pairs as representing a continuum of ratios of the expertise of A to B (with a lot of redundant pairs). So the last line would represent A as being very much an expert compared to B. Then as we move down the last row, B 'catches up' with A in expertise step by step. So expertise is clearly not a binary notion. And, of course, expertise is not a discrete notion. It is truly continuous so instead of the matrix represented by the set of ordered pairs above, there is an infinite array of possibilities. If we want to try to imagine this array, the best we can do is look at an illustration of the diagonal argument that the rational numbers are countable.



And of course this leaves out infinite infinities of irrational numbers that might express the ratio of A's expertise to B's. The weight of A's testimony might be 6.262 times that of B (or so we may suppose, it doesn't affect the point). I have gone to pains with the visuals in order to illustrate how, well, infinitely short a binary model of expertise would fall from the mark.

Someone who *is* an authority (to some degree) in this way *has* authority (to some degree) in this way. And since this kind of authority – the kind an expert has (we might as well call it 'expertise') – is based on various good-making features of their beliefs,[5] we can sensibly call it

[5] It wouldn't have to be knowledge, and often likely isn't. What's really important is that their beliefs be sufficiently likely to be true, on some relevant notion(s) of 'likely'.

epistemic authority. There is nothing at all mysterious about this kind of epistemic authority. We meet it in the doctor's office, in the classroom, at church, at work, and many other places in daily life.

The Lockean thesis that faith is belief on the basis of the authority[6] of the proposer, when 'authority' is used to express expert authority (as I believe Locke intended it) is a sensible, familiar idea. It explains why 'walking by faith, not sight' can be reasonable. I've never seen the Great Wall of China. But someone who has been there many times can tell me something about it, say that it is wider in inner Mongolia than it is down in Beijing, and my (reasonably) believing his testimony. That is, we are (epistemically) trusting her (like we would a thermometer). We are having (epistemic)[7] faith that though we have not seen it with our own eyes, we yet believe that it is so. Note that it is an open question whether this kind of belief counts as knowledge. Sometimes, it seems clear enough, that belief on the basis of expert testimony counts as knowledge. Other times, it will be less clear. But on this notion of faith some of the things we believe by faith we know by faith. The matter of first importance, though, is that belief on the basis of epistemic authority is perfectly reasonable. In non-binary terms, we are warranted in having great confidence in many items we take by faith. We can sensibly assign them high degrees of probability. This is the notion of epistemic authority I find in Locke and find to be true. My position is that this notion of epistemic authority is the best model for 'faith that' in religious belief. Not that on my account trusting an authority that p is just a species of believing that p. Trusting (epistemically) in some authority is just using them as a source of justification. And someone needs evidence that someone is an expert to reasonably treat them as an authority. This is very much in the spirit of Hume as well. For this reason, and because it is a very unpopular term, I'll call the view I'm advocating here the 'Enlightenment' view. It is important to note here that I do not accept the standard narrative according to which Locke and Hume removed all communal notions from their epistemology, which Zagzebski seems to place a lot of stock in (p. 112). However, I am very glad we share this common ground: 'I think Locke is right that faith is tied to belief on

[6] Really it is upon the *perceived* authority, but that is merely a wrinkle.

[7] Epistemic faith I take to be a species of ordinary belief. I affirm the tradition of distinguishing between 'faith that' (what I have called 'epistemic faith') and 'faith in' which is a kind of interpersonal trust of individuals. 'Saving faith' or 'trusting in God/Christ' is a species of 'faith in'. I am not treating 'faith in' in this essay.

testimony, and the rationality of faith is therefore tied to the rationality of belief on testimony.' (p. 112)

The obvious sense of the Enlightenment view makes it hard to understand Zagzebski's final words in her chapter about religious epistemic authority: 'Religious faith is impossible to explain, much less justify, on the evidence view of testimony. That view forces us to either redefine faith as belief on a certain kind of evidence, as Locke did, or we must say that faith is nonrational.' (p. 179) That faith (that) is a kind of belief on a kind of evidence is run of the mill, as the examples above show. So there is no need for any redefinition. Things are fine just as they are. And the following is also unfair to Locke:

> Taking a belief from Scripture or from a religious tradition would not be justified at all according to the extreme egoist. That is to say, the belief would not be justified because it is from Scripture or tradition. If the belief is justified, it is because it is justified by the use of my faculties anyway. The fact that the same belief is included in Scripture or the tradition is irrelevant. (p. 167)

What's true is that it would not be justified *just because* it is from Scripture or tradition, for there are many scriptures and many traditions, and one must have some reason – independent of the *mere* fact that it is Scripture G or Tradition T – to believe G or T on the matter rather than X or Y. But from this it does not follow that it is *irrelevant* that it is included in Scripture or tradition for the simple reason that had Scripture or tradition not included it I never would have known it (or not as easily or as clearly), since my faculties are by themselves insufficient to attain to many of the truths revealed in Scripture and tradition. Also, it's because Scripture and tradition bear a certain relation to God Himself that they are worthy of belief. So on any reasonable view at all one does not believe something *just because* it is from Scripture or tradition. Below, I will defend Locke's view on authority and choice.

1.2 Juridical Authority

The other kind of authority I will call *juridical* authority. This is the authority you might think a Sergeant in the Army has over a Private, a CEO has over an office manager, that deans seem to think they have over professors, and that Police would have over Citizens if a government were just. I'm dubious of the notion of juridical authority as such. I'm inclined to think that as moral agents no one can have authority over

us – including God – unless we grant it to them. But I don't need to defend that here.

'What is essential to authority', says Zagzebski, 'is that it is a normative power that generates reasons for others to do or to believe something preemptively' (p. 89). The key term is 'preemptively'. She notes the presence of pre-emption in Joseph Raz's notion of juridical authority. '[T]he fact that an authority requires performance of an action is a reason for its performance that *replaces* other relevant reasons and is not simply added to them' (p. 93, emphasis added). Then she presents her epistemic analogue:

> *Preemption Thesis for epistemic authority*
> The fact that the authority has a belief p is a reason for me to believe p that replaces my other reasons relevant to believing p and is not simply added to them.

She observes that this principle might be thought to be impossible: We must ignore our own reasons. But she says we do not ignore our reasons when we pre-empt them. 'In fact, it is because I am not ignoring them that I see that the belief of the authority has a certain status vis-à-vis my other reasons.' (p. 98) It is all well and good to *say* this, but what is utterly mysterious is how we can focus on our reasons (the opposite of ignoring them) and yet they have no motive force (apart from malfunction). How can we look a reason 'square in the face' and remain (non-degenerately) unmoved? There are deviant cases, of course, but this is supposed to happen systematically and virtuously. She relies on a dubious difference between first-person reasons and third-person reasons, saying that they do not aggregate (pp. 56-58). But why not? Why can't reasons of different kinds aggregate? They seem to in ordinary cases of testimony (where we don't have to think of the testifier as an authority). I see grocery bags on the floor and conclude that Jim when to the market. Jill tells me he did. I rightly believe more firmly when Jill's testimony is added to my experience. Very different kinds of reasons aggregate. Furthermore, her distinction rests on the distinction between first- and third-person reasons, but I reject the notion that there are any third-person reasons but rather that all reasons are first-personal (see Dougherty and Rysiew 2009, Conee and Feldman 2004, Heumer 2001, Swinburne 2001). It is this dubious alleged difference between first- and third-person reasons which supports her reasoning in her chapter on religious faith that the evidence theory of testimony is false (p. 168). A word about that is in order.

Zagzebski adopts the common sentiment that there is more to testimony than evidence. 'As Moran (2005) has pointed out, there is no explanation for feeling let down when the testimony is false if testimony is evidence.' (p. 108) Of course there is more to testimony than evidence. It's a human practice, so there is bound to be a social normative dimension. The question is whether this 'more' plays any role in the justification of beliefs obtained from testimony. And the social stuff itself is part of what *makes* testimony evidence-generating. I can tell, instinctively, when someone is representing themselves as having a certain degree of warrant for a proposition, p, expressed by an utterance they make. I am attending, often unconsciously, to features of the social situation *which themselves are of evidential significance*. Grice's rules of interpretation are psycho-social in nature, and the textbook Gricean interlocutor uses social norms as premises in an argument for the conclusion that S has certain information relative to p.

Zagzebski says (p. 118) the 'egoist' or proponent of the Enlightenment view should or can accept her main theses regarding testimony. (p. 116)

> *Justification Thesis 1 for the Authority of Testimony (JAT 1)*
> The authority of a person's testimony for me is justified by my conscientious judgment that I am more likely to satisfy my desire to get true beliefs and avoid false beliefs if I believe what the authority tells me than if I try to figure out what to believe myself.
>
> *Justification Thesis 2 for the Authority of Testimony (JAT 2)*
> The authority of another person's testimony for me is justified by my conscientious judgment that if I believe what the authority tells me, the result will survive my conscientious self-reflection better than if I try to figure out what to believe myself.

When I first read these principles I was perplexed as to how they depended on any of the prior material. So I was a bit relieved when she said that Lockeans could accept them or modified versions of them. However, I can't quite tell if this is correct, because she says this:

> Suppose I hear speaker S say that p and reasonably take that to be either direct or inductive evidence that p . I then believe p on S's testimony. If I also have evidence that S is more likely to get the truth whether p than I am, then I ought to believe p on her testimony, and I ought to believe preemptively. (p. 118)

Because of the weighting model I presented above, I just don't think there is any need at all for preemption. That is, I see no reason *at all* to just set aside my own reasoning *per se*. There are only two scenarios where I can see this as reasonable. First, there is the case where I have thought it through but gain evidence sufficient to make *certain* that I have absolutely *no idea* what I'm doing. In this case, I should abandon any hope in my reasoning and give it no weight. But if I am less than certain that there is no hope, then I ought to give my reasoning *some* weight (and, as is clear, the maths allows me to give it arbitrarily low weight). The other scenario isn't technically of the same kind, that is it is not a case of setting aside reasoning at all, but it is the analogue in the realm of action. This is the scenario where I am considering whether to investigate a matter and think it through. Before I am to begin I learn that you are in so much better an epistemic position than me that it would be a waste of my time and resources to even bother about it. The odds of my investigation and reasoning affecting my outcome credence by a significant margin are so low that thinking about it myself has negative expected utility. In this case I 'set aside' my *attempt* to think through the matter myself. This can make perfect sense because I haven't yet gone to the effort and therefore have no reasons to 'set aside'. Outside of these two cases – the one merely theoretical, since I don't think one should ever assign 1 or 0 to any proposition – and the other in the realm of action – I cannot think of any reason to ever utterly discount our own reasoning. Unlike the former case, the latter may be quite widespread, and I hypothesize that this kind of phenomenon is what is driving the bus, and it is just a confusion to think of it as relating to epistemic reasons. The reasons in this case are practical reasons.

There is much left out of this characterization which is important, especially Zagzebski's notion of the communal in religious authority. However, that material is built upon the distinction between theoretical and deliberative reasons and, especially, on the notion of epistemic authority as *pre-emptive*, wherein one *sets aside* one's own reasoning. My arguments below will be aimed primarily at the pre-emptiveness of juridical epistemic authority.

2.1 THE ARGUMENT FROM PARSIMONY

My first argument in favour of the Enlightenment view over the Juridical or Pre-emption view is an argument from parsimony.

Parsimony argument

(1) The notion of expert epistemic authority is completely non-mysterious and one both sides are committed to.

(2) Expert authority is sufficient to explain the relevant epistemic practices (esp. the reasonableness of accepting testimony).

(3) If 1 and 2, then unless there is compelling reason to posit a new, mysterious notion of epistemic authority, one should not do so.

(4) There is no compelling reason to posit a new, mysterious notion of epistemic authority.

Therefore,

(5) There we should not posit a new, mysterious notion of epistemic authority.

Premise 1 seems secure. Premise 3 is an expression of the Principle of Parsimony: don't posit new things without due cause. Partial defence of 2 and 4 come from remarks above.

2.2 The argument from the aim of belief

The second argument begins from the familiar notion that belief, in some sense or other, 'aims' at truth. I actually think belief aims at empirical adequacy (Dougherty 2014), but I suspect that Zagzebski and many others hold some kind of view that will do the trick.

B1 Belief aims at truth.

B2 If belief aims at truth, then if a mental state doesn't aim at truth, that state isn't one of belief. [B1]

B3 If a mental state doesn't aim at truth, that state isn't one of belief. [B1, B2 MP]

B4 Any state that is formed (wholly) in response to considerations other than signs of truth (i.e. evidence, broadly construed) is not aimed at truth.

B5 Any state that is formed (wholly) in response to considerations other than evidence, isn't one of belief. [B3, B4 HS]

B6 The state of assent formed in response to authority is not a response to evidential considerations (else it wouldn't really be epistemic *authority*).

B7 The state of assent formed in response to authority is not a state of belief. [B5, B6 HS]

All of the basic premises of this argument seem true by definition to me, so I'm unsure what to even try to defend (or how I would do it).

2.3 Degrees of Authority

Epistemic authority, as Zagzebski conceives it, is binary in a certain way. It appears that you should set aside your own reasoning when you rely on an epistemic authority in her sense. It is not, as it would be with expert authority in my sense, combined and weighed with your own judgment. She says, 'a small difference between myself and the putative authority is not likely to be sufficient to ground authority' (p. 96). The idea seems to be that there needs to be a large difference between one and a putative authority before one pre-empts one's own reasoning. Presumably, prior to crossing this critical threshold one continues to rely on one's own reasoning. But this seems strange. It ignores the difference between small differences in expertise and not-so-small differences in expertise.[8]

On the generalization of testimonial evidence I have provided (Dougherty 2013), weighing one's own reasoning against the epistemic authority is as easy as calculating a student's final grade. Say Maria is an expert with respect to some class K of propositions to which p belongs (and is the only relevant category in this case), and she is .8 confident that p. It seems to me, let us suppose, that she is being too cautious. My own judgment is that on the evidence available to Maria and me, one should rate p at .9. However, I recognize that Maria is an expert to the following degree: her opinion should count twice as much as mine. My final credence in p then is figured thusly: (.8 + .8 + .9)/3 = .833. The method is the same even if there is only 'a small difference' between Maria and me. If her opinion should count for 110% of mine[9] we have (.9 + 1.1(.8))/2.1 = .848. As expected, this is just slightly below the straight (unweighted) average of our credences, .85. This is a fully general

[8] There is also the problem of the vagueness itself, which I don't have space to go into in any detail here. The problem is that there will be no bright line such that from exactly that line on we pre-empt our own reasoning and below it we don't. But we must either pre-empt or not pre-empt. So there seems to be a necessary mismatch between reality and practice on the pre-emption model.

[9] This is actually consonant with Zagzebski's 'Principle of Epistemic Trust in Others' (p. 186), and she does talk about degrees of trust (p. 188), but then she goes back to binary language when talking about resolution of disagreement: 'Given the argument of this book, it is reasonable to resolve the conflict in favor of what I trust the most when I am thinking in the way I trust the most' (p. 189) and that resolving disagreements means we need to 'target the belief that should be given up' (p. 191).

account of how to weigh authority. Furthermore, this method can handle multiple competing authorities of various weights. It is just like figuring a slightly more complex grade with more quizzes, a midterm and a final, all of different weights. And this is a very good result, since epistemic authority clearly comes in degrees. It is hard to see how Zagzebski's account can handle degrees of epistemic authority in such an elegant way or even at all. 'Why isn't it more reasonable', she asks, 'to add my other reasons to the balance of reasons, perhaps weighing the authority's belief more heavily than my other reasons? Isn't the authority's belief just one more piece of evidence that I put into the mix of my total evidence?' (p. 99) She favours pre-emption over weighing. This seems like exactly the wrong result in light of the existence of degrees of authority.

III. REJOINDER TO ZAGZEBSKI ON BEHALF OF LOCKE

Zagzebski considers Locke's objection to a duty to believe in obedience to authority stemming from the fact that belief does not depend upon the will. In objection, she offers essentially this argument.

N1 Beliefs have norms.
N2 If beliefs have norms, then we can exercise control over them.

Therefore,

N3 We can exercise control over beliefs.

In defence of N1 she says that we teach norms to students (p. 87). It's not clear what norms she is referring to. Here is the sort of norm you might think she has in mind (to a student who is writing a research paper, say): *Base your thesis on adequate research*. But this comes to little more than the conjunctive injunction *Do research and have it in mind when writing*. But this isn't a norm of belief at all. This is a norm of writing. We might get more specific with *When you write down your thesis, think about what evidence there is for it amidst your research*. That is at least an instruction to think about something. Though this is not an injunction to form a belief, it can be expected to result in beliefs for a properly functioning person. So I'm going to ask you right now to think about when your birthday is. The odds are, being an agreeable sort of reader, you complied. And odds are if you did think about it you spontaneously formed the belief that your birthday was on such and such a day. The formation of the belief itself was spontaneous, it was not up to you to do or not to do. There

is likely nothing you could do to resist it. So the belief was unfree, not something you were responsible for. Yet it was something over which you had some kind of positive control. For any given number within a certain limit, you can quickly comply with a request to form the belief that it is not prime. You do this not by taking direct aim at the formation of a belief but rather by directly bringing it about that you are thinking about either some evidence for some target proposition or by simply thinking of the content of some self-evident proposition.

The above fits will the term 'reflective self-control', which Zagzebski sometimes uses to describe the kind of control she thinks we have over belief. But it is hard to see what application this could have for the idea of epistemic authority in the juridical sense (vs. the expert sense). For if the belief is formed in virtue of one's responsiveness to ordinary reasons, then there is no role played by juridical authority. So even if one was being obedient to juridical authority in engaging in the act of reflection, that isn't the salient explanation of the formation of the belief. At least the dominant explanation is evidential. Someone, in an attempt to get me to believe Goldbach's conjecture, might order me to go into a room. Once in the room I see Goldbach's conjecture written on the wall quite large and come to believe it. We would hardly highlight the juridical authority of the one who sent me to my room as the dominant *epistemic* explanation of my belief in Goldbach's conjecture. The juridical authority merely *occasioned it.* So though there is a certain kind of guidance control over belief, it is not of the sort to provide a model for pre-empting reasons.

CONCLUSION

My position is that God's authority is the authority of the expert, like the expert on horses, who has no jurisdiction at all, whereas one who does have jurisdiction may lack authority due to ignorance (the latter claim is not strictly necessary for my argument). There is not much the Pope can actually do to me, but has 'moral authority' because he is a reliable source of information about what to do in life generally. (And I have also made a vow of obedience to the Church when I was confirmed as a Catholic, but I freely made that vow, so the juridical authority it gives rise to is not a problem for my view.) We are wrong to model God's authority (or the Church's authority, which ultimately derives from God's authority) on that of juridical authority. The reason to obey God in general is not that he is the 'Big Boss' with so much power,

but rather because as a loving God he wants what is best for us and as an omniscient God he always knows what that is. And so it is, Lockeans say, with belief. We should believe what God says or the Church says not because of a position they have over us but rather because we have good reason to believe that God and the Church want the best for us and God always and the Church usually knows what that is (the Church within its defined sphere).[10]

BIBLIOGRAPHY

Conee, Earl & Feldman, Richard. 2004. *Evidentialism* (Oxford: Oxford University Press)

Dougherty, Trent. 2014. 'The Ethics of Belief is Ethics (Period)', in Tim O'Connor and Lara Goins (eds), *Faith and Epistemic Virtue* (Oxford: Oxford University Press)

Dougherty, Trent. 2013. 'Dealing with Disagreement from the First Person Perspective: A Probabilist Proposal', in Diego Machuca (ed.), *Disagreement and Skepticism*, Routledge Studies in Contemporary Philosophy, vol. 46 (New York: Routledge)

Dougherty, Trent & Rysiew, Patrick. 2009. 'Fallibilism, Epistemic Possibility, and Concessive Knowledge Attributions', *Philosophy and Phenomenological Research*, 78 (1): 123-132

Feldman, Richard. 2004. 'Having Evidence', in Conee and Feldman (2004).

Huemer, Michael. 2001. *Skepticism and the Veil of Perception* (Lanham: Rowman & Littlefield)

Leland, John. 1740. *Answer to a Book Entitled 'Christianity as Old as the Creation' 2nd ed.* (London: Richard Hett)

Meadley, G. W. 1809. *Memoirs of William Paley, DD* (London: J. Graham)

Swinburne, Richard. 2001. *Epistemic Justification* (Oxford: Oxford University Press)

[10] This essay was made possible through the support of a grant from the Templeton Religion Trust. The opinions expressed in this publication are those of the author and do not necessarily reflect the views of the Templeton Religion Trust.

ZAGZEBSKI ON AUTHORITY AND PRE-EMPTION IN THE DOMAIN OF BELIEF

ARNON KEREN

University of Haifa

Linda Zagzebski's *Epistemic Authority: A Theory of Trust, Authority, and Autonomy of Belief* is an important contribution to the study of epistemology, and social epistemology in particular. It is also a very timely contribution. For too long, epistemologists and social epistemologists have talked about authority in the domain of belief without paying enough attention to the very concept of epistemic authority. Zagzebski's book is bound to change this.

Zagzebski presents a detailed account of the concept of epistemic authority, describing the essential features of authority which allow us to talk of authority both in the practical realm and in the realm of belief. Building on Joseph Raz's account of political and practical authority, she claims that to have authority is to have a special kind of normative power: 'a normative power that generates reasons for others to do or to believe something preemptively.' (Zagzebski 2012: 102) Thus she has directed the philosophical community's attention to an important concept that has previously received only little attention: to the concept of pre-emptive reasons for belief, a reason for belief that 'replaces other reasons the subject has' (Zagzebski 2012: 102), rather than being added to them.[1] She has thus done more than anyone else to bring to the attention of the philosophical community the two central concepts,

[1] Raz himself has *in passim* addressed the applicability of his account of practical authority to the domain of belief. See, e.g., Raz (1986: 29-30, 52-3); (2006: 1032-7). In Keren (2006; 2007) and elsewhere, I have studied the concept of epistemic authority, and the applicability of the Razian account to it. Like Zagzebski, I have defended the applicability of the notion of pre-emptive reason to the domain of belief and its centrality to our understanding of epistemic authority. Nonetheless, Zagzebski's recent account differs from my own, and my discussion here will touch upon some of the differences.

of epistemic authority and of pre-emptive reasons for belief, and the relation between them.

With some qualifications, I think that Zagzebski's central claims about the conceptual relations between these two concepts are correct. I also agree with her central normative claim that we are often justified in believing on authority, and are sometimes rationally required to do so. Indeed, I have defended similar conceptual and normative claims in my Ph.D. dissertation (Keren 2006) and in subsequent writings (Keren 2007; 2014). Nonetheless, I think that in arguing for these claims, Zagzebski has ignored important differences between practical authority and authority in the domain of belief, and that as a result her attempt to explain why it is rational to believe on authority is lacking. Because I think that these differences are of much epistemological significance, my discussion here will focus on Zagzebski's attempt to show that Raz's account of political authority, when properly interpreted, is applicable to the domain of beliefs (Zagzebski 2012: ch. 5), and will try to show where her application of the Razian account goes wrong. Obviously, this will leave untouched much of what Zagzebski does in this rich and important book.

Raz develops his account of authority by claiming that a number of theses – content independence, the pre-emption-, dependence-, and normal justification theses – are true of political authorities, and that some – e.g., the no-difference thesis – are not. Zagzebski attempts to show that analogues of the theses that are satisfied by political authorities are also satisfied by epistemic authorities; and that the no-difference thesis, which is not true of political authorities, is also not true of epistemic authorities. Thus, the basic contours of epistemic authority match the contours that Raz ascribes to political authority.

This, I argue, is only partly true. Starting with the conceptual claim about what epistemic authority and believing on authority consist in, I argue (§1) that Zagzebski is correct in identifying the pre-emptive nature of reasons provided by an authority as central to our understanding of epistemic authority. Thus Zagzebski is correct that Raz's pre-emption thesis is satisfied by epistemic authority. However, while this central feature is shared by epistemic and political authorities, other significant features are not. Thus, I argue (§2) that the no-difference thesis, when interpreted in the way intended by Raz, is true of epistemic authorities, despite being false of political authorities, and that Zagzebski's characterization of the distinguishing feature of authority can therefore

be misleading. Finally, I turn to the rationality of believing on authority (§3): I argue that Zagzebski's attempt to show that an analogue of the normal justification thesis applies to the domain of belief again ignores key differences between belief and action, and that her explanation of the rationality of believing on authority therefore fails. A successful explanation of this will need to be more attuned to the differences between political and epistemic authorities.

I. AUTHORITY, TRUST, AND PRE-EMPTION

Zagzebski opens her book by noting that authority in the domain of belief, unlike practical authority, receives little attention in contemporary philosophy. Her aim is to change this, and to convince us of the rationality of believing on authority. Focusing mainly on the point of view of the subject, of a subject asking herself 'how she should get beliefs she accepts upon reflection', she claims that 'we are all committed to accepting epistemic authority' (Zagzebski 2012: 2-3). Her argument for this claim is built of two main stages. In the first, she argues that epistemic-trust in others is not rationally escapable. By virtue of our rational, inescapable trust in ourselves, and by virtue of the similarity between ourselves and others, we are committed to placing epistemic-trust in others.[2] At the second stage she attempts to establish that 'among those we are committed to trusting are some whom we ought to treat as epistemic authorities' (Zagzebski 2012: 3).

As noted, the second stage of Zagzebski's argument builds heavily – at times too heavily, I shall argue – on Joseph Raz's account of political authority. Zagzebski adopts two main Razian themes: one conceptual, one normative. The first involves claims about the conceptual relations between epistemic authority and the normative power to generate pre-emptive reasons for belief, and the centrality of the latter for our understating of what believing or acting on authority consists in. The normative claim is that it is rational for us to believe on authority, and that what justifies belief on authority parallels what, according to Raz, justifies acting on someone's authority.

Let us start by considering the conceptual question. Zagzebski does not seem to present an argument for the claim that believing on authority

[2] A second feature of self trust that commits us to epistemically trusting others is our rational trust in our emotions, and in particular, in our epistemic admiration of others (Zagzebski 2012: 93).

essentially involves taking the authority's belief as issuing a pre-emptive reason for belief. Nonetheless, there are reasons to think that she is correct about this. One type of consideration supporting this claim is based on the kind normative responses that are open to us when we invite others to treat us as epistemic authorities, or when we advise others to form beliefs on a person's authority. If to have epistemic authority is to have the normative power to generate pre-emptive reasons for belief, then there should be distinct forms of criticism that will be open to us when we think that a thinker should treat another as an epistemic authority: her refusal to let the authority's belief replace her own consideration of the evidence will then count as grounds for criticizing her. Indeed, this is a form of criticism we are apt to make when we think that a subject should believe upon others' authority, e.g., when we think that laypersons should defer to the authority of scientists but they refuse to do so,[3] or when children refuse to defer to the epistemic authority of their parents. Correlatively, forms of criticism that are often available to us should not be available to us when we think that a subject should treat another as an epistemic authority: if we point out to person *A* that she should believe that *p* on *B*'s epistemic authority, then we cannot at the same time suggest that *A* should weigh all the relevant evidence available to her, or criticize her for not doing so. Indeed, such criticism of a subject does seem to be incompatible with the suggestion that he should believe on another thinker's authority.

Such considerations lend support to Zagzebski's claim about the conceptual relations between having epistemic authority, and having the power to generate pre-emptive reasons for belief. However, they may also suggest that her characterization of the distinguishing feature of epistemic authority in these terms should be slightly amended. For there are reasons to think that being epistemically trustworthy also involves the power to issue pre-emptive reasons for belief, and that one can be thus trustworthy on an issue without having epistemic authority on it.

Thus, consider a speaker, Trevor, who address us and says: 'Trust me, *p*.' And suppose that we do form the belief that *p*, and do so without considering evidence available to us which is relevant to the truth of *p*. Trevor, it seems, cannot then criticize us for not considering all the evidence available to us. Such criticism seems to be incompatible with his invitation to us that we trust him. If he were to criticize us in this

[3] See e.g. Jones (2002).

way, he would seem to be withdrawing his original appeal to us that we trust him (Keren 2014). Compare this with the case of Esther, who intentionally provides us with very good evidence for *p*, expecting us to recognize this intention of hers. If here we also form the belief that *p* without weighing other available evidence, Esther, it seems, can criticize us by saying: 'You are correct. But you should have considered all the evidence before reaching that conclusion.' Such criticism, whether valid or not, seems to be perfectly compatible with her expectation of us that we recognize that she intends to provide us with very good evidence, but not with Trevor's expectation that we trust him.

Thus, there are forms of criticism that are often available to us, but that are not compatible with inviting someone to trust, or with advising someone to form a belief upon epistemic trust. And these are the very forms of criticism that we should expect to be closed, if pre-emption is central not only to believing on authority, but also to believing on trust. The point is not merely that sometimes a trustworthy person's belief can provide us with pre-emptive reasons. The point is a conceptual point about trust and trustworthiness. Trusting a person, quite generally, requires seeing yourself as having, in virtue of the person's trustworthiness, reason against taking certain precautions, and thus as having pre-emptive reasons for action or belief (Keren 2014):[4] You do not trust the babysitter, if you don't think that in virtue of her trustworthiness, you have reason not to install, as a precaution, nanny cameras throughout the house; you do not trust a thinker for the truth, if you don't think that in virtue of her trustworthiness, you have reason not to weigh all evidence available to you, just as a precaution.

This might require introducing a minor revision in Zagzebski's account of the distinguishing features of epistemic authority. For we might want to distinguish, as Zagzebski does, between being epistemically trustworthy and having epistemic authority, and between believing on trust, and believing on authority.[5] Even if our notion of trust essentially involves the idea of pre-emption, there is a distinction to be

[4] Reasons against taking precautions are reasons against acting for precautionary reasons. Hence they are reasons against acting for certain other reasons, and hence, pre-emptive reasons. See Keren (2014).

[5] Zagzebski assumes that we can believe upon trust without believing on authority because she employs a notion of epistemic trust weaker than the one singled-out here, and which entails seeing a trusted thinker's belief that *p* as providing us with a prima-facie reason for believing *p*, but not with a pre-emptive reason for believing *p*.

made between the kind of pre-emptive reasons for belief provided by the report of an authoritative expert, and the pre-emptive reasons for belief typically provided to the expert by a trustworthy layperson's report. Even if the expert trusts the layperson, she may not treat her as an authority. But then it is not the ability to generate pre-emptive reasons for belief that distinguishes having epistemic authority from being epistemically trustworthy. What distinguishes between the two is the force of the pre-emptive reasons that one has the power to generate. The belief of epistemic authorities provides me with pre-emptive reasons for belief that make it epistemically irresponsible for me to form the relevant belief on my own weighing of the evidence. In contrast, the beliefs of trustworthy thinkers give me reasons that need not render it epistemically irresponsible for me to form the belief on my own weighing of the evidence, even if they also make it responsible for me to allow the expert's belief to pre-empt my own. In this sense, what is distinctive of epistemic authority is that it is the normative power not just to provide others with reasons to believe something pre-emptively, but that it is the power to generate to others an epistemic *duty* to believe something pre-emptively.[6]

II. GENERATING PRE-EMPTIVE REASONS FOR BELIEF

Believing on authority thus involves seeing ourselves as having a pre-emptive reason for belief. There is thus a sense in which Zagzebski is correct that one has epistemic authority only if one has the normative power to generate such reasons. However, there is an important sense of 'generating reasons' in which this is false: a sense in which practical authorities indeed have the normative power to generate reasons to do things pre-emptively, but in which epistemic authorities do not have the normative power to generate pre-emptive reasons for belief. This is arguably an important aspect of a proper understanding of the working of pre-emptive reasons for belief and of the very concept of epistemic authority.

In presenting his account of practical authority, Raz discusses, and rejects, what he calls the no-difference thesis, according to which, 'the

[6] In this sense, the kind of normative power distinctive of epistemic authorities is more similar to the normative power of practical and political authorities than Zagzebski's formulation suggests. To say that I have (practical) authority over my son is to say not only that I can generate reasons for him to do things pre-emptively, but that I can generate duties for my son (Enoch 2014).

exercise of authority should make no difference to what its subject ought to do for it ought to direct them to do what they ought to do in any event' (Raz 1986: 48). Raz rightly rejects this thesis when applied to practical authorities: an authoritative directive stating that we must all drive on the left-hand side of the street can create for us a duty to drive on the left-hand side of the street, where we wouldn't have such a duty otherwise.

Zagzebski claims that the epistemic analogue of the no-difference thesis is also false. If the analogue is that 'the fact that an epistemic authority believes a certain proposition *p* should make no difference to what I ought to believe', as Zagzebski suggests (2012: 109), then she is probably correct about that. Epistemic authorities do generate reasons in the weak sense that knowing what they believe often makes a difference to what we ought to believe. But such a reading of the no-difference thesis is not the one intended by Raz. In discussing the thesis, Raz clearly has in mind a stronger sense of 'generating reasons'. For the no-difference thesis which Raz rejects does not deny that authoritative pronouncements have epistemic significance; instead, it denies that authoritative pronouncements have normative significance beyond their epistemic significance. According to the thesis, '[t]here is nothing which those subject to authority ought to do as a result of the exercise of authority which they did not have to do independently of the exercise, they *merely have new reasons for believing* that certain acts were prohibited or obligatory all along' (Raz 1986: 30; emphasis mine). This is arguably false of practical authorities, for an authoritative command can create for us a good reasons to Φ (to drive on the left-hand of the street), in the sense that had the command not been given, no one would have had a reason, let alone a duty, to Φ; nonetheless, the reasons to Φ generated by the command is a perfectly good reason. In contrast, an authoritative speaker's testimony or belief that *p* cannot generate for us reasons to believe that *p* in this strong sense. If the authoritative thinker did not have good reasons to believe that *p* herself, then the reason to believe that *p* generated by her testimony that *p* is not a good reason. It is a misleading reason.

Therefore, to have epistemic authority, is to have the normative power to generate for others reasons to believe something pre-emptively, but not in the sense in which to have practical authority is to have the power to generate for others reasons to do something pre-emptively. Practical authorities can, but epistemic authorities cannot, generate reasons in the strong sense which figures in Raz's discussion. An epistemic authority

cannot provide us with good reasons for belief if she does not have one herself; she can merely give us good reasons to believe *p* if there were good reasons to believe *p* all along. It might therefore be more accurate to say that epistemic authority consists in the power to pre-emptively *transfer* to others good reasons for belief. Or better: to provide others with pre-emptive reasons for belief that allow the authority's good reasons for believing something to support others' belief in the same thing.

Importantly, this strong sense of 'generating reasons', which figures in Raz's discussion, and which distinguishes between epistemic and practical authority, is one that should interest epistemologists interested, like Zagzebski, in figuring out how consideration of authority figures in the reflective subject's own perspective. When we believe that *p* on a speaker's authority, this is only because we believe that she herself has, independently of her telling us that *p,* a good reason to believe that *p.* More broadly, it is this strong sense of 'generating reasons' that should interest epistemologists in their attempt to understand how, and under what conditions, we can obtain knowledge by trusting other thinkers. Thus an appreciation of the sense in which epistemic authority cannot generate reasons to believe allows us to make better sense of the controversy over the principle stating that testimony does not generate knowledge: the principle stating that a hearer can come to know that *p* on the basis of a speaker's testimony that *p* only if the speaker herself knows that *p*. In recent years, a number of apparent counter-examples have emerged to this principle: cases in which a hearer appears to come to know that *p* on the basis of the testimony of a speaker who does not know that *p* (Graham 2000). Armed with an understanding of the sense in which epistemic authorities cannot generate reasons pre-emptively, we can understand in what sense this principle is correct, in spite of apparent counter-examples; we can understand why it is right to dismiss these counterexamples as cases where knowledge is based on testimony but not in the normal way (Williamson 2000: 257). To trust a speaker, we have noted, is to see her testimony as providing us with pre-emptive reasons to believe what she says; but epistemic authorities, and epistemically trustworthy speakers more generally, can generate pre-emptive reasons for belief only in the weak sense. Thus the reason made available to audiences by the speaker's testimony, in as much as it is made available to them through the distinctive normative structure associated with trust, cannot be stronger than the reason available to the speaker herself. In as much as a belief owes its epistemic status to

trust, its epistemic status cannot be better than the epistemic status of the belief of the trusted person. If a hearer comes to know by believing an unknowing speaker's testimony, this can only be because her testimonially-based belief is supported by other reasons beyond the reasons for belief provided by the testimony itself, reasons which were not pre-empted by the speaker's testimony. Therefore, the justification for the hearer's belief is not a product of the distinctive normative structure associated with epistemic trust and authority, and her way of gaining knowledge is not the normal way of obtaining knowledge through testimony (Keren 2007).

III. JUSTIFIED BELIEF ON AUTHORITY

The characterization of the distinguishing features of epistemic authority leaves open the question whether we should ever form a belief on a person's authority: Should we ever treat the fact that a putative authority believes that *p* not just as a reason for believing *p*, but as one that has pre-emptive force? Zagzebski defends a positive answer to the question, and does so by applying to the domain of belief another Razian thesis – the normal justification thesis (NJT).

According to NJT,

> the normal way to establish that a person has authority over another person involves showing that the alleged subject is likely better to comply with reasons which apply to him (other than the alleged authoritative directives) if he accepts the directives of the alleged authority as authoritatively binding and tries to follow them, rather than by trying to follow the reasons which apply to him directly. (Raz 1986: 53)

Zagzebski adopts an analogous thesis about the justification of epistemic authority:

> Justification Thesis 1 for the Authority of Belief (JAB 1): The authority of another person's belief for me is justified by my conscientious judgment that I am more likely to form a true belief and avoid a false belief if I believe what the authority believes than if I try to figure out what to believe myself. (Zagzebski 2012: 110)

She then employs a Razian argument – the *financial shares argument* (Raz 1986: 67-69) – to show that the conditions set by JAB1 are indeed satisfied by pre-emptively believing on authority.

There are thus two questions to ask about Zagzebski's NJT-based argument for the justification of believing on authority: (1) does JAB1 indeed set a correct standard for the justification of belief on authority? (2) Does Raz's financial shares argument, when applied to the case of belief, indeed show that believing on authority satisfies the conditions set by JAB1? It seems to me that Zagzebski's positive answer to both questions is mistaken. Zagzebski seems to ignore important differences between the practical and the epistemic domain, which undermine her application of the Razian justification of acting on authority to the justification of believing on authority.

Consider first the standard of justification set by JAB1. Both NJT and JAB1 assume a strong connection between doing the best one can, and being justified. NJT assumes that acting on authority is justified if and because doing so is the best way one has of complying with reasons that apply independently. Similarly, JAB1 assumes that believing on authority is justified if and because doing so is the best way one has of achieving the epistemic goals of believing truth and avoiding error. However, there are reasons for doubting the parallel between NJT and JAB1; even if NJT sets standards sufficient for justification in the practical domain, the same is not true of the standards set by JAB1 for the domain of belief. In the practical domain, there is a tight connection between what one can best do, and what one is justified in doing. Thus, very different moral theories can all agree that an agent ought to perform an act if and only if it is the best action that she has available (while disagreeing on what makes an action 'best').[7] However, this kind of intimate relation does not exist in the case of belief: that one can do no better than form a belief in a certain way does not mean that the belief is epistemically justified. If the only belief-forming mechanisms available to a thinker are unreliable and yield beliefs that do not fit the evidence, then the fact that a belief was formed by using the least unreliable belief-forming mechanism available does not mean that the belief is epistemically justified.

[7] See, e.g., Zimmerman (1996). Some views about supererogation might attempt to loosen this strong connection between 'ought' and 'good', to allow for the possibility of actions that are beyond the call of duty: thus they would allow that action *A* may be better than *B*, but that we have no duty to perform *A* and may permissibly perform *B*. But even such views preserve the following strong relation between 'best' and 'ought'. Even if one may perform an action which is not best, in the relevant sense of 'best' picked by the moral theory, it is always the case that one may also perform whatever action is best. No moral theory would suggest that an action which is best cannot justifiably be performed.

Accordingly, the known inferiority of other belief-forming strategies available to me does not suffice to show that I am justified in believing on the authority of another person. Suppose that ignorant Al knows that if he himself tries to figure out what to believe on a certain question, he is very likely to form a false belief, and that the likelihood of mistake will be slightly lower, but still very high, if he treats sophomore Sue's judgment as authoritative. This knowledge does not justify Al's belief formed by deferring to Sue's authority.

Indeed, if Al, while being confused on the matter at hand, is not confused on what epistemic authority consists in, he too will arguably not see Sue as an epistemic authority to whose judgment he ought to defer. Thus, when we believe that a person has authority on whether *p*, we can cite this fact, and the fact that she told me that *p*, to explain how we know that *p.* But if, having formed the belief that *p* by deferring to sophomore Sue, Al is asked how he knows that *p,* he cannot answer the challenge by pointing to the inferiority of other belief forming mechanisms available to him. This is just not the right kind of answer to the question: 'How do you know?'

Even if the kind of normative evaluation we are interested in involves a weak notion of epistemic entitlement, one that does not suffice to render true belief knowledge, it is doubtful whether JAB1 provides us with an adequate account. JAB1 appears inadequate even if we take it as an account of reasonable belief, or if we focus only on the reflective thinker's own perspective. Thus, it is unclear whether a subject can hold a belief, knowing that he only holds it because this is what an unreliable thinker believes, and because other available ways of deciding on what to believe are even less reliable. Again, this does not seem to be the right kind of reason when it comes to belief. Moreover, even if he could somehow sustain his belief while knowing that it is not supported by the right kind of reason, because he can know this, his belief will arguably not be a reasonable one, at least not in an epistemic sense.

To some extent, Zagzebski seems to be aware of this problem with JAB1. Thus she notes that JAB1 is 'not sufficient to justify taking a belief on epistemic authority without qualifications ... [for] I might judge that even though the putative authority is more likely to get the truth whether *p* than I, the authority is not very likely to get the truth either' (Zagzebski 2012: 111). However, she does not explain what other conditions must be met for a belief on authority to be justified or reasonable. But our discussion of what epistemic authority consists in may suggest what

element, missing from JAB1, must be part of any plausible justification of belief on authority: If to have authority is to have the power to provide others with pre-emptive reasons for belief that allow one's good reasons for believing something to support others' belief in the same thing, then to be justified in believing on authority, one must be justified in believing that the putative authority has good reasons for believing as she does. This is what Al is not justified in believing about Sue.

Zagzebski also fails to provide a good explanation of why the conditions specified by JAB1 are likely to be fulfilled by believing on authority: Why are we more likely to form a true belief and avoid a false belief if we treat a putative authority's belief as giving us a reason that pre-empts other evidence, rather than by adding it to other evidence available to us? If I do not trust my own judgment, then it may be obvious why I should let someone else's judgment replace my own. But this is not a line of argument available to Zagzebski, whose argument for the rationality of believing on authority is based on the rational inescapability of self-trust. And while the argument from self-trust might establish that sometimes I should trust another person's way of getting a belief more than I trust the way in which I would get a belief, it is not clear why Zagzebski thinks that this justifies believing on the person's authority. Why think that in such a case 'the conscientious thing to do is to let the other person stand in for me in my attempt to get the truth' (2012: 105)? After all, Zagzebski employs a weak notion of trust, such that her claim about the rational inescapability of trusting others only entails that the beliefs of those trusted gives me a *prima facie* reason to believe as they do (2012: 68). So if I have a reason to trust the putative authority's way of forming beliefs more than I trust my own, does this not require treating it as a weightier prima facie reason, rather than as a reason with pre-emptive force?

Zagzebski (2012: 114) responds to this challenge by drawing on Raz's financial shares argument. Consider a case where I am faced with a practical decision – 'whether or not to sell certain shares' (Raz 1986: 67), and suppose that I am given advice by a financial expert, such that all that I know about her is that she is more likely to make the right decision than I am when I form an independent judgment. Raz argues that in such a case I can do no better than by allowing the expert's judgment to pre-empt my judgment altogether. If instead I treat it as providing me with an additional *prima facie* reason, I will not do as well. For 'only by allowing the authority's judgment to pre-empt mine altogether will I succeed in improving my performance and bringing it to the level of

the authority' (1986: 68): Only in this way can I lower my rate of mistake so that it matches that of the expert. If instead I assign some weight to my own independent judgment and some (greater) weight to hers, in a way that will allow her judgment to reverse my own in a certain proportion of the cases, then my rate of mistake would still be higher than the expert's, even if it would be lower than the rate of mistake of my independent judgment: My rate of mistake will likely match the authority's low rate of mistake in that proportion of the cases where my judgment conforms to hers, but will match the higher rate of mistake of my independent judgment in those cases where my judgment differs from hers.

Zagzebski accepts this argument as sound, and takes it as showing that in such cases 'treating [an] authority's belief that *p* as just one reason among others to believe *p* ... will worsen my track record in getting the truth' (2012: 114). For if Raz's argument is sound, she claims, 'it does not matter whether the authority's judgment is about what to do or about what to believe' (Zagzebski 2012: 115). But this is a mistake. While I agree that under certain conditions taking beliefs on authority is both justified and the best epistemic strategy available to us, I don't think that Raz's financial shares argument shows this. For, *pace* Zagzebski, it *does* matter whether we are considering a practical question, of the kind considered by Raz, or the question what to believe.

In the kind of practical case described by Raz, there are only two qualitatively different options – 'Sell' and 'Don't Sell' – and two possible states of the world: one in which 'Sell' is best, another in which 'Don't Sell' is best. Given appropriate assumptions, which are arguably satisfied in the case described by Raz,[8] always following the expert's judgment makes it most likely that you will choose the best option (and most likely that you will avoid the worst), and is therefore the best possible strategy. However, in cases where the question is what to believe, even if there are also just two relevant states of the world – *p* and not-*p*, there are always (at least) three qualitatively different options: 'believe *p*', 'believe

[8] For Raz's argument to work, it does not suffice that always conforming our judgment to that of the expert will make us least likely to make a mistake, that is, to choose an option that is not best. To be successful, the argument must also make assumptions about how the payoffs of a mistake will depend on the actual state of the world. It will succeed, e.g., if the payoff of the best option is the same in the case where the best option is 'sell' and where the best option is 'don't sell' (and likewise for the worst option), so that we do not care more about making one type of mistake rather than the other. But this condition is often satisfied, and is arguably often satisfied in the kind of case discussed by Raz.

not-*p*', and 'suspend judgment'. In the practical case described by Raz, there is no third option, in the sense that the possible payoffs of not-deciding are identical to that of deciding not to sell. In contrast, in the case of belief the possible epistemic payoffs of withholding judgment are different both from those of 'believing *p*' and of 'believing not-*p*'. Even if the expert's rate of mistake is lower than my own independent rate of mistake, so that always following the expert's judgment makes it most likely that I will choose the best option (believing truth), it does not follow that this is the best possible strategy. For I can significantly improve my chances of avoiding the worst option – falsely believing – if I do not allow the expert's judgment to pre-empt my own. By suspending judgment in at least some proportion of the cases in which she believes *p* and I independently believe not-*p* (or vice versa) – cases in which she is much more likely to be mistaken compared to cases on which we agree – I can lower the probability of error not only below that of my own independent judgment, but also below that of the expert. Accordingly, even if all I know is that she is more likely to form the correct judgment than I independently am, when the issue is what to believe, it is simply not the case that I will be more likely to form a true belief and avoid a false belief if I allow the expert's judgment to pre-empt mine.[9]

I conclude therefore that Zagzebski's attempt to justify our believing on authority, by applying the Razian framework to the domain of belief, is inadequate. It fails to set an adequate standard for the justification of belief on authority; and it fails to show that believing on authority ever satisfies those standards that it does set. This does not mean that

[9] Note that this objection to the applicability of Raz's financial-shares argument to the case of belief depends on Zagzebski's specification of our epistemic goals in terms of believing truth and avoiding error. If these are our epistemic goals then we have an option – that of withholding judgment – that, while not best, allows us to guarantee an epistemic payoff which is second-best. It is because of the availability of this option that we should not always adopt the expert's judgment, even if doing so is most likely to result in the optimal payoff. If, however, our epistemic goal is different from that specified by Zagzebski, then it might be possible to apply Raz's financial-shares argument to the case of belief without succumbing to the objection, if this alternative conception does not allow us to guarantee an epistemic payoff which is second-best. This may suggest, in line with other objections to Zagzebski's account made here, that a successful justification of believing on authority would most likely appeal to the epistemic goal of holding doxastic attitudes that fit the evidence. Zagzebski comes close at one point to proposing that believing on authority should be justified in terms of such an evidential goal, but rejects this, claiming that believing on the evidence is neither our ultimate epistemic goal, nor our most important one (2012: 110).

Zagzebski is mistaken in her important claim, that we are often justified in treating a thinker's belief as a reason that has pre-emptive force and that we are sometimes rationally required to do so. On the contrary, this important claim seems to me to be quite true (Keren 2006; 2014). However, what our discussion suggests is that in order to show that it is true we must be more attuned to the differences between the way authority functions in the practical and epistemic domains.

Our discussion suggests that both an adequate account of what epistemic authority consists in, and an adequate explanation of why believing on authority can be justified, must appeal to the good reasons for the (same) belief that the authoritative thinker has. In accounting for what epistemic authority consists in, we ought not to say that it consists in the ability to generate reasons to believe something pre-emptively in the strong sense; instead, it consists in the power to provide others with pre-emptive reasons for belief that allow their beliefs to be supported by the authority's good reasons for belief. And in presenting an account of the justification of our believing on authority, we will have to appeal to these good reasons, and to the support they can provide to our belief: first, in specifying what conditions must be met for our belief on authority to be justified; and second, in explaining how, by allowing the authority's judgment to pre-empt our own, these conditions can be met. In this, the justification of believing on authority will differ not only in details, but in structure, from the justification of acting on authority. In explaining our justification for doing Φ on an authoritative command, we need not appeal to reasons for Φ'ing existing independently of the command. After all, it is the command itself that can generate these reasons. In contrast, in explaining our justification for believing p on an authoritative testimony, we must appeal to reasons for p which the authority has independently of her telling us that p, and to our reasons for believing that such reasons might support our own belief if we believe on authority.

Acknowledgement. Work on this paper was supported by the Israel Science Foundation (grant no. 714/12). For helpful comments on the paper, I am grateful to Ariel Meirav and Saul Smilansky. For their helpful comments, advice and encouragement throughout my work on the concept of epistemic authority, I am especially grateful to Philip Kitcher and Joseph Raz.

BIBLIOGRAPHY

Enoch, David. 2014. 'Authority and Reason-Giving', *Philosophy and Phenomenological Research,* 89 (2): 296-332

Graham, Peter. 2000. 'Conveying Information', *Synthese*, 123 (3): 365–92

Jones, Ward E. 2002. 'Dissident versus Loyalist: Which scientists should we trust?', *The Journal of Value Inquiry*, 36 (4): 511-20

Keren, Arnon. 2006. 'Testimony, Science and the Social: Obtaining Knowledge from Authorities', PhD diss., Columbia University

Keren, Arnon. 2007. 'Epistemic Authority, Testimony and the Transmission of Knowledge', *Episteme*, 4: 368-81

Keren, Arnon. 2014. 'Trust and Belief: A Pre-emptive Reasons Account', *Synthese,* 191 (12): 2593-615

Raz, Joseph. 1986. *The Morality of Freedom* (Oxford: Clarendon Press)

Raz, Joseph. 2006. 'The Problem of Authority: Revisiting the Service Conception', *Minnesota Law Review*, 90: 1003-44

Williamson, Timothy. 2000. *Knowledge and its Limits* (Oxford: Oxford University Press)

Zagzebski, Linda Trinkaus. 2012. *Epistemic Authority: A Theory of Trust, Authority, and Autonomy in Belief* (Oxford: Oxford University Press)

Zimmerman, Michael. 1996. *The Concept of Moral Obligation* (Cambridge: Cambridge University Press)

ZAGZEBSKI ON MODELS OF REVELATION

JACEK WOJTYSIAK

John Paul II Catholic University of Lublin

From the numerous themes present in a deep and inspiring book by Linda Zagzebski, *Epistemic Authority*, I would like to address some problems related to the model of revelation (chap. 9, sec. 3.2, pp. 191-199).[1] It is particularly important for theologians and followers of the religions, such as Judaism, Christianity and Islam, which 'have in common a revelation from God at a time in the past' (p. 192). Both theologians and ordinary believers want to understand the content of the revelation and guarantee its inviolable transmission through generations. A philosopher of religion, however, asks whether it is at all possible for such a revelation to occur, and if so, how it is to be recognized and in what ways it may be accessible 'to a great number of people over a very long period of time in widely varying circumstances' (p. 191). Zagzebski discusses three out of a variety of models of revelation conceived of as a specific kind of communication between God and the human being ('communication between God and me' – p. 191). Using her key words, let us call them, respectively: the chain model, the experience model, and the high point (or the state of perfection) model.

I. ASSUMPTIONS

Before I examine, in discussion with Linda Zagzebski, the mentioned models, I will reconstruct the assumptions that she accepted and that, consequently, affected her choice, presentation and evaluation of those models. (Zagzebski clearly sympathizes with the last one). In my opinion,

[1] If not indicated otherwise, all page numbers refer to the discussed book by Linda Trinkaus Zagzebski (2012).

those assumptions – and their equivalents presented and defended in various parts of her book – come down to the following:

> **The assumption of historicity:** God's revelation took place at a definite moment or period in the history of humankind and is transmitted or updated in the times that follow it. (This assumption does not exclude the existence of non-historical revelation, but leaves it out of the scope of analysis.)
>
> **The assumption of communality:** The recipient of revelation is a member of a definite community, of a certain *We*, within which a revelation takes place and is transmitted.
>
> **The assumption of the indispensability of authority:** No human being can acquire correct beliefs on *divine matters*, or lead a proper religious life, relying exclusively on herself, in a cognitive isolation (or independence) from beliefs and exemplars of life of the community that is a partner in a revelation.
>
> **The assumption of the test of conscientious judgment:** The condition of a given person's justified acceptance of a given religious community authority (as a partner or a bearer of a revelation) is the person's conscientious judgment that if she engages in the beliefs and practices of this community – instead of forming them on her own – the chances that their result will survive her conscientious self-reflection will increase.[2]

I believe that a good model of revelation should not only agree with the listed assumptions, but should also provide more specific criteria to recognize a revelation and to correctly choose a religious authority from competing candidates. As the author herself remarks, 'cases of competing authorities' (p. 111) and the fact 'of disagreement between communities' (p. 221), and especially between communal religious authorities, constitute one of the greatest challenges for epistemology (and especially for the epistemology of religion).

---

[2] The first assumption appears clearly on pp. 190-192, the three others on pp. 199-203. The assumption of communality seems a consequence of epistemic universalism, and the assumptions concerning authority are applications of the main argument of the book: in various areas of knowledge and life appear individuals or communities that are 'in a better position to get the truth [or other valuable ends] than I' (p. 111); if they pass the test of my conscientious self-reflection, 'I should follow the[ir] authority in that case' (p. 111).

II. THE CHAIN MODEL

I believe that the chain model of revelation (CM) may be represented – in agreement with Zagzebski's intentions (cf. pp. 193-194) – as a trichotomous structure:

(CM1) **initiating event**: God enters (at a definite historical time) in direct contact with the few chosen people (founders of religion, prophets, direct witnesses of divinity, etc.);

(CM2) **multistage transmission**: a result of the initiating event (in the form of stories, a set of rules, mystical poetry, etc.) 'is transmitted by oral or written testimony' to other people – first directly (eyewitnesses – their listeners or firsthand recipients), and later indirectly (firsthand recipients – secondhand recipients, etc.);

(CM3) **integrating intention**: in God's intention the initiating event (experience) contains important content that may be understood not only by an eyewitness, but also by all its *n*-hand recipients; due to this fact, all the recipients (regardless of when and where they live) can constitute one community – the community which, thanks to the acts of collective memory, continually makes available or updates the universal content of the initiating event.[3]

It seems obvious that the presented model agrees with the assumptions of historicity, communality and indispensability of authority. The only problem might at most be the fact of the privileged position of the initiating event participants. We should remember, however, that in the history of religions it is difficult to find isolated initiated events whose participants are deprived of the authoritative context of their own community.[4] For instance, Moses stood before God as a representative of his community with its tradition and mission for the future. The membership of this community enabled him to identify the Revealing One as the *God of his fathers*, and His commands as the commands

[3] Point (CM3) is necessary to answer questions like: 'Why would it matter to us what a man called Abraham did, or that Moses had a religious experience in front of a burning bush if we are only the distant recipients of testimony about their contact with God?' (p. 193)

[4] We should rather speak of various types (conditioned by historical-communal contexts and connected to one another) of initiating events that differentiate the families of religions (e.g. the family of the Abrahamic religions), religions (e.g. Christianity), denominations (e.g. Methodism), and intra-confessional spiritual traditions (e.g. Franciscan spirituality in Catholicism).

for the whole of *Israel*. Similar was the case of the Apostles who had initially interpreted the person of Jesus Christ in the context of the community of Israel, extending later its range under the influence of a new initiating event.

Let us assume the chain model and ask what more specific conditions must a given religious community, presumed to have been constituted by (CM1)-(CM3), meet to deserve a positive epistemic evaluation (according to the principle of the conscientious judgment test). In other words: When are we allowed to regard a given community as a reliable bearer of God's revelation or as a justified religious authority?

I believe that a person who attempts to judge the epistemic value of a given religious community should answer three questions:

(i) the question concerning (CM3): Does the community under evaluation proclaim contents comprehensible to its contemporaries and reveal universal truths about God and man?

(ii) the question concerning (CM2): Does the message transmitted by the community under evaluation accurately inform about the initiating event?

(iii) the question concerning (CM1): Has the initiating event (recounted in the message) actually been inspired by God?

I believe that a sympathetic interpretation of the contents proclaimed by the great world religions would allow an affirmative answer to the first question in almost all of the cases.[5] In such a situation, the evaluation of a given religious community would depend on the answers to the remaining questions.

As for the second question, the advocates of the community under evaluation might refer to the so-called *Expansion of Trust Principle*:

> 'I have reason to trust those who are conscientiously trusted by those I conscientiously trust.' (p. 97)

Indeed, in many religions their fundamental contents are transmitted on the basis of natural trust between numerous transmitters and recipients, conceived of as links in a multi-element chain. This trust is usually related to family ties: parents transmit their faith to children, the children to

[5] A less sympathetic or less open interpreter might judge that content from the vantage point of her prior (especially pre-religious) convictions concerning God, the world and values. Would such a person, however, need revelation, if she had known beforehand what exactly it should contain?

their children, etc. Sometimes faith is acquired by people from outside these ties. This, however, is often related either to trust based on the emotion of admiration (cf. pp. 87-93), or to a rational analysis of the presumed authority's cognitive capacity. In each of these cases – as we can say, paraphrasing the Expansion of Authority Principles (p. 152) – a person may be justified in believing that she has the proper access to the content of the initiating event, because she is justified in believing that it is not her, but the community or certain of its representatives who may have knowledge on this matter, and this community or its representatives are justified in believing that it is not them, but the first witness or witnesses (related to them) who have such knowledge. The access to the contents of the initiating event (experience) would then be guaranteed by the whole community as a chain of witnesses, initiated by the first witnesses. The links of this chain are bound together by the appropriate – most frequently based on trust or related emotions – acts of authority recognition.

Religious communities usually present themselves as faithful transmitters of the content of the initiating event. Their members have *prima facie* good reasons to believe in the message transmitted by their community. (Whom is the person born in this community to believe, if not her ancestors? And the one who joined the community basing on her own judgment – whom is she to believe, if not her ability to recognize authority?) The question arises, however, whether the community's message on the initiating event is sufficient to accept the claim to its divine origin. In this way we approach the problem (CM1) – the first and most important element of the chain model.

An adherent of the affirmative answer to the above question reasons more or less in the following way:

> If I accept the testimony (coming from the first witnesses) of a given religious community on the initiating event, I should also accept its interpretation of this event (coming, explicitly or implicitly, from the first witnesses), including the interpretation that points to its divine origin; since the competences of the community (greatly exceeding my own) concern not only testimony transmission, but also its interpretation.

Let us call this approach the internal approach (IA), as it occurs most frequently in those who reason, as it were, from inside of a given

community. Their adversaries – we will call them adherents of the external approach (EA) – might reason in the following way:

> A religious community (and its first witnesses) is capable only of the transmission of information on the initiating event, but not of its interpretation; in that case I am allowed to accept the divine origin of the event only if I reach the belief that a description of this event (provided by the community and the first witnesses) contains such elements that if I had been its direct participant, I would have accepted (immediately or on further reflection) on the basis of those elements that the event had been inspired by God.

(EA) is most frequently used in religious apologetics. Not only do apologists wish to present the self-understanding of a given religious community, but, in the first place, they wish to present reasons that might induce a person from outside the community to accept its claims to divine origin. Among such reasons, Christian apologists give a special place to miracles. They usually claim that the trustworthy Christian message on the initiating events of Christianity includes information on miracles that confirm the divine origin of those events. For instance, Richard Swinburne (2008: 85-87) emphasizes that 'we need evidence of God's "signature" on the prophet's work', and that its key element is 'a violation of laws of nature'. Moreover, 'the particular violation must be of a kind which the culture in which the violation occurred would recognize as God's signature.' For Christians, such a sign would be the Resurrection of Christ presented in the trustworthy reports of the New Testament.[6]

I do not intend to evaluate here the claims of Christianity (or any other religion) to the authority in matters of divine revelation. The above analyses aimed only at showing that within the chain model it is possible to reconstruct how a given person may be justified in her belief about the access to divine revelation: this person may have good reasons to accept that a given community – through a chain of testimonies – reliably recounts the initiating event, which is of divine and universal character.

What would happen, however, when the mentioned person came across counter-reasons that oppose her belief? Those counter-reasons may be essentially reduced to the arguments supporting the thesis

[6] Cf. William L. Craig (2008: 333): 'If Jesus rose from the dead, then his claims are vindicated and our Christian hope is sure; if Jesus did not rise, our faith is futile [...]'. Cf. 1 Corinthians 15: 14-15.

that – as Linda Zagzebski writes (p. 202) – reports on the so called initiating events 'have probably been distorted in multiple ways during the course of many centuries'. But how do we know this? An adherent of this thesis might refer either to the scientific knowledge concerning the times when the initiating events allegedly occurred, or point to the discrepancies between the relevant testimonies. Let us observe, however, that scientific knowledge concerning the past largely relies on the records of human testimonies on it.[7] Therefore, the testimony of a community actually competes not with the pure knowledge of the past, but with other testimonies. All is thus reduced to the competition of testimonies. Every one of us, in various situations of life, must evaluate on one's own the reliability of the witnesses (sometimes competing) she encounters. This concerns also the religious sphere. It may happen that competing testimonies will induce an adherent of a given religion to abandon it or to convert to another one, or to correct her beliefs. However, it may also happen that those testimonies will be, in her opinion, too weak to change her beliefs, or that the testimony of her own community is more important to her. In the latter case she will continue believing in her original religion.[8]

III. THE EXPERIENCE MODEL

As Linda Zagzebski observes, the experience model – as opposed to the chain model – 'focuses on the recipient's experience rather than the original experience [(CM1)]' and reduces or 'minimizes the function of tradition [(CM2)] in preserving the past intact' (p. 194). In other words, in this model the nature of religious life does not consist in collective remembering of the remote initiating event, but amounts to the '*first*hand experience of the divine' (p. 194). This experience is shared by all the

[7] My analysis omits the knowledge interpreted in the light of the principle of naturalism (claiming that no supernatural events can take place), because making such an assumption one cannot envisage any possibility of revelation.

[8] Let us remember that discrepancies between testimonies do not always concern essential issues. Besides, religious communities, especially those sharing common roots, dispute their interpretations of the message rather than its literal wording. It cannot also be excluded that the interpretations are complementary, and their multiplicity has been intended by the Revealing One who allows different ways of His revelation – according to the different cognitive capacities and spiritual needs of the revelation recipients. This issue merits a separate discussion, for instance, in the context of Zagzebski's pertinent example of the disagreement concerning the Holy Trinity (pp. 213, 219-221).

confessors, and not only by God's chosen ones who, in the past, came into a direct contact with Him.

Then, how does it happen that individual religious persons, each of whom experiences God in the way limited 'to the experience of one person' (p. 202), form a community where a divine revelation acquires intersubjective validity? I suppose that it is so, because the experience model incorporates the principles of the chain model, giving them new functions. If this is true, one can say that in the experience model:

> (EM1) the initiating event does not consist so much in an exclusive and immediate experience of God by the few people He elected, as it provides inspiration for all future experiences of God by many more people; it may also be the norm that enables evaluation whether a given experience is a proper experience of God;
>
> (EM2) the function of the multistage transmission of the result of the initiating event (experience) does not consist so much in its faithful preservation for future generations, as in enabling still new experiences similar to the original one;[9]
>
> (EM3) the integrating intention of God (the Holy Spirit) unites all believers into one community, but not on the principle of collective remembering, but thanks to the common content of their experiences, which in some way imitate the original experience.

As we can see, the above – and I believe the optimum – interpretation of the experience model presents it as a reformulation and expansion of the chain model. In consequence, the presently discussed model inherits the advantages of the previous model. Moreover it has an advantage absent in the chain model: it is the possibility of subjects' having additional reasons to accept a given religion, i.e. the subject can accept it not only on the basis of her conviction about the community's reliability in transmitting the testimony of the original revelatory event, but also on the basis of her personal experience of the universal contents of this event. The fact of this experience being shared by many people in different times and places makes it impossible to raise an objection of subjectivism and isolationism.

However, what will happen if a subject of an experience typical for a given religion meets a person with a radically different religious

[9] As Zagzebski writes (p. 197): in Christianity 'the transmission of the Gospels is the occasion by which the Holy Spirit produces faith as a first-hand relation to God'.

experience? Then, like in the case of the competition of testimonies, the competition of experiences will occur. The difference between the two competitions consists in that, in the former case, the subject evaluates testimonies of others, while in the latter, she compares her own experiences to those of others. She can hardly be expected to give priority to someone else's experience over her own, to which she has a privileged first-person access. With William P. Alston (1993: 274), she might say:

> 'In the absence of any external reason for supposing that one of the competing [epistemic] practices [or experiences] is more accurate than my own, the only rational course for me is to sit tight with the practice of which I am a master and which serves me so well in guiding my activity in the world' and in the religion.

Finishing the analysis of the experience model, I would like to add that that some Christians emphasize the importance of the mentioned experience so strongly that they feel exempt from the epistemic concern for reasons to confirm the community's reliability in transmitting the testimony. I consider this an epistemic error. Without recognizing this reliability we cannot be justified in believing that our experiences are similar to the normative original experiences, and that the unity of different confessors' experiences is neither illusory nor accidental. Moreover, extreme interpretations of the experience model lead to the change in the concept of revelation, whose content – as I have already said – significantly includes (or ought to include) the factors of historicity, communality and authority.

IV. THE HIGH POINT MODEL

The last model identifies divine revelation neither with the original experience (that is transmitted further), nor with subsequent experiences (inspired or regulated by the original one). In this model, revelation is identified (as Zagzebski puts it, p. 195, 197) with a certain 'way of understanding God': a given community acquired such a way of understanding in the period that it recognized as 'a high point' in its history. This way of understanding God is expressed in a collection – considered as exemplary – of (oral and written) beliefs, stories, prayers, rites, practices, laws, commands, etc. We may say that – according to the model in question – God reveals Himself not so much in individual events or experiences, as in a self-understanding of a given religious

community at the time when it was in a 'state of perfection' or mastery in the fulfilment of its activities.[10]

The merit of the high point model is its capacity, so that it can incorporate the previous models. Thus the mentioned exemplar of faith and practice of a given community may also include the story concerning the initiating event and 'the belief that the chain [of its transmission] is basically accurate' (p. 202) and unbroken. The exemplar may also include descriptions of paradigmatic religious experiences or other elements inspiring such experiences. Those factors are indeed very important for the life of religious communities; communities, however, transmit them in a wider context – in the context of once acquired, rich and compound wisdom.

An additional advantage of the discussed model is that it allows a wider range of reasons available to the subject who wishes to justify her acceptance of a given religion. The subject may thus refer not only to the reliability of a given community in transmitting the testimony of the initiating event or to her own or someone else's religious experiences, but also – or above all – to the cognitive and existential importance of the exemplar of beliefs and practices cultivated by the community. This is the case when the subject passes the conscientious judgment expressive of the following (or similar) content:

> If I accept the exemplar of beliefs and practices proposed by a given religious community (instead of relying on another community or on myself only), my present and future convictions, feelings, experiences, actions, etc., will be coherently organized, justified or oriented, and 'will satisfy my future conscientious self-reflection' (p. 199).

The above judgement may be based either on the recognition of the quasi-moral characteristics of the very community (such as its longevity; the number of people who accepted its teachings and were not disappointed; multifarious fruitfulness, e.g. charitable or culture-creating, of its activities), or on the recognition of the epistemic value of beliefs it proclaims. In the latter case the judgment would be a kind of religious hypothesis, which – as the Polish logician Józef Bocheński

[10] As examples Zagzebski gives 'the end of the biblical era' for Judaism and 'the Apostolic age' (and early post-Apostolic tradition) for (Catholic) Christianity. A secular analogy is the artisanal mastery achieved in 16th century by Venetian glassmakers whose technique has been imitated in Italian glassmaking until today (see p. 196).

suggested (1965: 148) – plays in the religious life a role similar to that of a scientific hypothesis in reductive sciences.[11]

The above-presented advantage of the high point model cannot – in my opinion – undermine its double dependence on the chain model. The first dependence consists in that the exemplary collection of a community's beliefs (formed at its high point in the past) should necessarily, and not just facultatively, include a reliable story on the initiating event as a particular God's intervention, confirmed by special signs. Without His intervention, the emergence of this community as one inspired by God would be difficult to understand. Without it the community could offer its members or candidates nothing more than a system of human life wisdom – a system that would appear merely as one of the options on the market of competing worldviews.

The second dependence is conceptual. It is impossible to describe the essence of the high point model without using the principles of the chain model, as the former has also a trichotomous structure that results from the modification of the original model. The modification consists in that:

(HM1) the concept of the initiating event has been replaced by the concept of exemplary 'way of understanding God', formed at the time of a community's flourishing;

(HM2) the multistage transmission, mentioned in (CM2), refers now to the transmission not of the initiating event itself, but of the whole 'way of understanding God' as an exemplary system of beliefs and practices;

(HM3) the integrating (community-creating) intention of God is fulfilled not only by recalling the remote event, or producing or imitating similar experiences, but by applying the exemplar of faith

[11] Bocheński gave particular attention to two (fallible in their nature) ways of justifying religious beliefs: through the authority and through the religious hypothesis. In his theory of authority, he emphasized (among others) that every human authority is relativized to a given field, i.e. 'every authority [except God] is only an authority in a limited class of sentences, not for all sentences' (1965: 172). If so, one cannot suspect a cognitive dissonance in a person who accepts the authority of a given individual or community in some religious issues, and questions it in others. In turn the concept of the religious hypothesis is clarified by him in the following way: 'at a certain time of his life the subject begins to think that, if he does accept the BD [basic dogma] of a certain religion, then the whole of his experience will become organized and somewhat explained' (1965: 149).

> and practice to varying conditions of every believer's life – so that she can achieve 'a way of living contact with God' (pp. 197-198).

Let us observe that the above presentation of the high point model – as a structure based on (CM1)-(CM3) structure – emphasizes the necessary components of revelation mentioned earlier: historicity, communality and indispensable authority. However, the specific principles (HM1)-(HM3) refer to different authorities or different layers of one authority: the authority of God, the authority of the exemplar of faith and of the experience of God (together with the authority of the initiating event transmission and the authority of its participants), the authority of the whole community, and the authority of institutions to which the 'responsibility for the authenticity [and updating] of tradition' (p. 198) has been ascribed. Examining the epistemic value of the authority conceived of in this way, it is necessary to consider more factors than in the case of less comprehensive models. All this renders the question of evaluating the revelatory claims, made by different religions, particularly complicated.

CONCLUSION

My analyses confirm Linda Zagzebski's belief that in looking for an adequate model of revelation, we need to go beyond the traditional chain model, as well as the model that attaches too much attention to experience. The high point model (or – as I would prefer to call it – the exemplary wisdom model) that she reconstructs actually seems to provide the most accurate description of the functioning of religion. Its greatest advantage is a very capacious understanding of revelation and authority. This capacity translates into a very wide range of reasons available to a subject to justify her choice of a certain religion (and also a wide range of problems she must consider in this context). It should be remembered, however, that the above model depends on the chain model in a significant way. In my opinion, the main epistemic concern of a person assessing the value of a given religion usually is – and should be – whether a given community speaks about facts and whether those facts consisted in the (continually updated) action of God.

BIBLIOGRAPHY

Alston, William P. 1993. *Perceiving God: The Epistemology of Religious Experience* (Ithaca and London: Cornell University Press)

Bochenski, Joseph M. 1965. *The Logic of Religion* (New York: New York University Press)

Craig, William Lane. 2008. *Reasonable Faith: Christian Truth and Apologetics* (Wheaton, Illinois: Crossway Books)

Swinburne, Richard. 2008. *Was Jesus God?* (Oxford: Oxford University Press)

Zagzebski, Linda Trinkaus. 2012. *Epistemic Authority: A Theory of Trust, Authority, and Autonomy in Belief* (Oxford: Oxford University Press)

EPISTEMIC AUTHORITY AND CONSCIENTIOUS BELIEF

CHARITY ANDERSON

University of Oxford

In *Epistemic Authority: A Theory of Trust, Authority, and Autonomy in Belief*[1], Linda Zagzebski engages in a wide-reaching investigation of underexplored epistemic terrain. The work is timely, given the recent interest of contemporary epistemologists in testimony and trust, and she approaches the topic from a novel angle: the rationality of beliefs taken on the authority of others. This accessibly written book covers an extensive span of topics including trust, testimony, and authority, in addition to the intersection of these notions with the domains of emotion, morality, religion, communities, and disagreement. Given the breadth of topics in the book, I will not be able to address all the ideas worthy of discussion. I will concentrate on a few topics. I will first examine the notions that serve as the foundation for the book: rationality, conscientiousness, and trust. I then raise several questions concerning her account of epistemic authority. Finally, I advance a worry regarding the application of her account to the problem of disagreement.

I. RATIONALITY, TRUST, AND CONSCIENTIOUSNESS

A main thesis of Zagzebski's book is that trust in oneself and in others is rationally inescapable. The self-reflective rationally consistent agent will trust others. Zagzebski takes as her opponent the epistemic egoist – someone who believes that her only way of gaining reasons for belief is to rely on her own faculties. The extreme epistemic egoist trusts only in herself she does not trust others (a less extreme egoist trusts others but only when her own faculties provide her with reasons to trust the person).

[1] Linda Trinkaus Zagzebski, *Epistemic Authority: A Theory of Trust, Authority, and Autonomy in Belief*, Oxford: Oxford University Press, 2012.

After defending the rationality of trust in both self and others, Zagzebski argues that it is rational to take beliefs on the authority of others, and then extends her argument to include the authority of communities as well as moral and religious authorities.

Rationality is central to Zagzebski's project. To be rational, she suggests is 'to do a better job of what we do in any case – what our faculties do naturally' (2012: 30). Our initial model of rationality, on her picture, is the resolution of the experience of dissonance or conflict between our mental states. Resolving dissonance is often automatic, but at times involves conscious deliberation. For Zagzebski, self-trust is rational because it resolves dissonance. More specifically, a general trust in one's faculties is rational because it resolves the dissonance arising from the epistemic circularity objection to justified belief in the reliability of our faculties. Particular trust in our faculties is rational when we are conscientious. Zagzebski understands epistemic conscientiousness as the 'quality of using our faculties to the best of our ability in order to get the truth' (2012: 48). It is the state or disposition of being careful and doing our best. Conscientiousness comes in degrees and requires self-awareness and self-monitoring. It is essential to her overall account that we trust we will succeed more often when we are conscientious. In this way, conscientiousness is our second model of rationality. As we shall see, this notion plays a key role in Zagzebski's defence of trusting epistemic authorities.

Zagzebski extends the rationality of trust in oneself to trust in others. Although she offers no clear set of conditions for when trust in others is justified, her main idea is that when we conscientiously judge that we can trust someone, we are rational to trust them. Zagzebski argues that if we are rational, we cannot fail to trust others, on pain of inconsistency. Her reasoning is as follows:

> When I am conscientious I will come to believe that other normal, mature humans have the same natural desire for truth and the same general powers and capacities that I have. If I have a general trust in myself and I accept the principle that I should treat like cases alike, I am rationally committed to having a general trust in them also. (2012: 55)

She takes it as given that we are all basically alike in our general abilities, and while she admits there are exceptions to this general similarity, she suggests the exceptions will be very limited. The principle of 'treat like alike' she understands as *a priori*.

I worry that this reasoning will apply in far fewer cases than Zagzebski seems to expect. Although most people may have the same general powers and capacities with respect to telling the time or knowing whether it is raining by looking out the window, in less mundane cases our powers and faculties vary in initial aptitude as well as specific training. If a rational requirement to trust others with respect to mundane facts is all that the argument establishes, this severely limits the extent to which trusting others is rationally inescapable and thereby reduces the effectiveness of the argument to oppose the egoist.

Zagzebski's reasoning leads her to affirm the following: the fact that another person believes *p* always gives me *prima facie* reason to believe *p* (2012: 58). At first glance this suggestion may seem suspect, but she goes on to explain that by 'reason' she does not mean a *decisive* reason, but rather one with 'genuine weight'. The fact that another person believes *p* counts in favour of *p* for me, though it does not always provide enough weight for me to believe *p*. Furthermore, the weight it provides can be defeated – in some cases it can be defeated quite easily. Given how weakly the principle is intended, the reasoning strikes me as plausible. If not read in light of this qualification, some of Zagzebski's statements about *prima facie* reasons could be misleading. One would be forgiven for thinking that she holds that one gains a sufficient reason to believe *p* when one discovers that someone believes *p*. In fact she doesn't think this. Nevertheless, Zagzebski maintains that it is *often* the case that trusting another person provides one with a decisive reason to believe what they believe.

Zagzebski extends the argument for general trust in others to particular trust in the faculties of others. When I discover that another person conscientiously believes *p* this provides me with a stronger *prima facie* reason to believe *p* (stronger than if I merely know the person believes *p*). It appears that she thinks most people exhibit conscientiousness most of the time and to roughly the same degree. She claims: '[w]hen I am believing conscientiously, I come to believe that many [others] are just as conscientious as I am when I am as conscientious as I can be.' (2012: 57) She further claims that when I am observant, 'I will inevitably have excellent evidence of the conscientiousness of others.' (2012: 61) Here we find a potential dilemma: being conscientious is either pretty easy, such that everyone is basically conscientious most of the time, or it is fairly demanding. If it is easy, it will not be able to play the role in solving the problems Zagzebski claims it solves. Or, at least, the solution will be

rather thin. Trusting our conscientious judgments is Zagzebski's advice for resolving various difficult epistemic situations, such as disagreement with those we trust. If we are generally conscientious all the time, then her responses to such problems will amount to the suggestion that we merely continue doing what we are already doing and generally always do. But if conscientiousness is demanding, the assumption that so many people exhibit it so often strikes me as unlikely. Most people do not continuously engage in self-reflection. We clearly do not use our faculties to the *best* of our ability all or even most of the time. But if conscientiousness is rare, then we won't readily have evidence of others' conscientiousness.

It's worth noting that Zagzebski seems to assume that we will know when we are conscientious – that when we are doing our best it will be evident that we are doing our best. This assumption strikes me as mistaken. Consider a hard-working student, eager to make good grades. He prepares for exams and tries to do his best, but while he often performs well he only sometimes does his best. It is easy to imagine that it is not obvious to him when he is doing his best and when he is not – at times he may be agnostic about his efforts until he receives his scores. Access to our own conscientiousness is further attenuated by our proneness to self-deception. When we fail we often console ourselves with the assurance that we did our best. The desire for this assurance provides a motivation to believe that we are doing our best most of the time. Our beliefs about our own conscientiousness are prime ground for self-deception.

One final issue arises in this connection. On Zagzebski's account, 'our only test that a belief is true is that it survives future conscientious reflection' (2012: 50). I ought to trust others when I predict their beliefs will satisfy my future self-reflection. In this way, being rational requires predicting the future in unrealistic ways. Zagzebski does not elaborate on how we determine which beliefs will satisfy future self-reflection, though she doesn't seem to expect that we will have trouble making such predictions. Given that we are not usually in a position to predict what our future evidence will be, it's hard to see how we could make judgments about future reflection with any confidence or accuracy.

II. EPISTEMIC AUTHORITIES

Zagzebski's defence of trusting epistemic authorities is the centrepiece of the book. An epistemic authority, on her view, is 'someone who does

what I would do if I were more conscientious or better than I am at satisfying the aim of conscientiousness – getting the truth' (2012: 109). The following principle, the Justification Thesis 2 for the Authority of Belief, states Zagzebski's main idea:

> (JAB2): The authority of another person's belief for me is justified by my conscientious judgment that I am more likely to form a belief that survives conscientious self-reflection if I believe what the authority believes than if I try to figure out what to believe myself. (2012: 110)

Trusting an authority differs from merely trusting another person. Trusting another person provides one with a *prima facie* reason to believe *p* (a reason that merely counts in favour of *p*), but authorities provide sufficient reasons to believe *p,* and further, authorities provide *pre-emptive* reasons.

Before I discuss pre-emption, I want to raise one question regarding who counts as an epistemic authority. It's not clear whether an epistemic authority must be more conscientious than me or whether someone could be an epistemic authority merely in virtue of having more evidence than me. Zagzebski's characterization of an epistemic authority suggests that an authority needs to be more conscientious than myself. An authority has 'more of the qualities I trust in myself insofar as I am epistemically conscientious' (2012: 108); she is someone who 'can help me to believe as I would believe myself, given my desires, emotions, and other beliefs ... '(2012: 111). Moreover, 'the point of epistemic authority is to help me in believing conscientiously' (2012: 111). In any case, these remarks indicate that an epistemic authority is not a general ideal, but a better version of yourself – when an epistemic authority is more conscientious than you, she must be conscientious in a way that resembles you. It seems I cannot rationally take someone as an epistemic authority if her psychic structure is incongruent with my current desires, emotions, and other beliefs. (If an epistemic authority must be more conscientious than myself, this sets a fairly stringent constraint on who can be an epistemic authority for me. Surely some of my desires and beliefs prevent me from being more conscientious than I am, thus the more conscientious version of me will lack those desires and beliefs though it's unclear how I would recognize which beliefs and desires those are in advance.)

In her account of what is *authoritative* about an epistemic authority, Zagzebski draws on contemporary work in legal philosophy – the writing of Joseph Raz in particular. Although in the practical realm authority is

typically exercised through commands, Zagzebski claims that the right to command is not an essential feature of authority. (She does mention that she thinks we are psychologically able to believe on command, but this supposition is not part of her account.) On Zagzebski's picture, what is essential to authority is that authorities generate *pre-emptive* reasons for a subject to act or to believe something. A pre-emptive reason is a 'reason that replaces other reasons the subject has' (2012: 102). The fact that an authority believes *p* is not simply added to one's other reasons; trusting an authority involves allowing the authority to 'stand in' for oneself. (It's worth noting one way in which beliefs taken on epistemic authority differ from those based on testimony: it can be rational to believe *p* on someone's authority merely when I know that that authority believes *p*. She does not have to tell me *p*; in fact, the authority does not even need to know that I exist.)

One concern Zagzebski raises concerning the idea of pre-emptive reasons is that it's not obvious that it is psychologically possible for me to let an authority's belief that *p* pre-empt my reasons. One might worry that I can't just ignore my other reasons. In defence of pre-emptive reasons, Zagzebski considers an example of a pre-emptive reason to act. She suggests that it is possible to stop at a red light for the reason that it is against the law to drive through a red light, even if one has other reasons for stopping, such as the likelihood of being hit. She states, 'if I stop because the law says to do so, that reason has the status of being my reason for stopping. It can be the reason even though I am quite capable of reciting many reasons for and against stopping.' (2012: 113) It appears that for the fact that the law requires one to stop at a red light to be my reason for stopping – and thus to pre-empt my other reasons – it must be the case that my other reasons do not play a role in motivating me to stop. That the law requires me to stop must be my basis for action. The parallel with belief is as follows: to believe *p* on authority the fact that the authority believes *p* must be my basis for believing.

I am sympathetic to Zagzebski's proposal that we sometimes make one of our reasons our sole basis for action or belief. But I worry that the settings where we adopt pre-emptive reasons – and thus the settings where we believe on authority – will be more limited than Zagzebski seems to envision. There are several reasons for this: first, because pre-emption is not always psychologically possible. Consider again a case of action. Suppose an authority commands me to take my mother to lunch on her birthday – something I was already planning to do because

I enjoy spending time with her and want her to enjoy her birthday. It's not clear that I can take her to lunch solely for the reason that the authority commands it – my other reasons are too strong to fail to motivate me. In the case of reasons for beliefs, the situation seems worse. If I already believe that it is wrong to steal, and have good reasons for so believing, I may not be able to believe solely because an authority believes it. My other reasons will inevitably play a role in my belief. The same will hold for many beliefs for which I have strong evidence.

Second, if pre-emption is essential for believing on authority, as Zagzebski suggests, it is not clear that it is possible to believe *p* on authority in a setting where I don't already have some reasons to believe or disbelieve *p.* Pre-emption seems to require prior possession of reasons – an authority's belief can't replace my reasons if I don't have any reasons to be replaced. (Perhaps we should understand a pre-emptive reason as a reason that replaces your other reasons *if you have any*.) Moreover, Zagzebski claims that when we already have beliefs on a topic, it is less likely we will trust an authority, since it is less likely we will judge that the authority's belief will survive our conscientious reflection. The foregoing considerations suggest that believing on authority will occur primarily in settings where either (1) I have no reasons to believe or disbelieve *p,* or, (2) I have weak reasons for or against *p*.

One final question regards whether pre-emption requires a kind of counterfactual stability. What if I let a reason pre-empt my other reasons but I very easily might not have? Suppose I stop at the red light because the law requires it, but if I had been in a hurry and no one was around I would have gone through the light. Is this fact relevant to the settings where I stop when I am not in a hurry? We can assume that when I am not in a hurry I believe that I am stopping because it is the law. It's not clear whether what I would do in similar cases makes a difference to whether a reason counts as pre-emptive. This issue may reach beyond the scope of Zagzebski's aims in this chapter, but I raise it as an area deserving further exploration.

III. CONSCIENTIOUS TRUST AND DISAGREEMENT

We've seen thus far that the notion of a belief that survives conscientious self-reflection plays a key role in Zagzebski's discussion. This theme is repeated throughout the work and is the foundation of the guidance she offers for navigating various epistemic difficulties, such as what one

ought to do in the face of disagreement. Zagzebski casts the problem of disagreement as an intrapersonal conflict, rather than an interpersonal conflict. On her view, disagreement is a problem that arises within self-trust because trusting myself commits me to trusting others. When I disagree with someone I trust, I experience conflict between my self-trust and my trust in the other person.

The appropriate resolution to disagreement with people I take to be conscientious, on her picture, is to believe in accordance with whatever 'I trust the most when I am thinking in the way I trust most, that is conscientiously' (2012: 214). When we are conscientious, we are doing our best and that is all we can do. Zagzebski provides surprisingly little guidance for dealing with cases of disagreement beyond this. She does suggest that beliefs that are central to one's noetic structure will often have survived a great deal of self-reflection; thus, when disagreement involves one of these core beliefs it is unlikely that one will judge that one ought to trust the other person. This way of thinking seems to condone a 'stick to your guns' mentality for at least many of our strongly held beliefs.

Resolution of disagreement involves prediction of what will survive future conscientious reflection:

> Disagreement with people we conscientiously judge to be conscientious should be handled ... in a way that we conscientiously judge will survive conscientious self-reflection. ... What is relevant for me is what I conscientiously believe, and what I predict will satisfy my future self-reflection, given what I conscientiously predict about myself. (2012: 215)

We have already mentioned one difficulty for this methodology: it requires predicting the future. A further question concerns what you ought to do if you know that you are bad at predicting what will satisfy your future self-reflection. Suppose you have loads of evidence that you are fairly unreliable at making such predictions. (Actually this strikes me as quite plausible, given that we often cannot predict what evidence will be available to us in the future and which beliefs will survive future self-reflection will be highly dependent on my future evidence.) Given what she's said thus far, it seems likely Zagzebki's response would be to do your conscientious best – that's all you can do.

It's a bit surprising that having dismissed the usual characterization of the problem of disagreement as not very interesting and failing to constitute a distinct evidential problem – despite the abundance of recent

puzzlement on the topic in the literature – Zagzebski's central advice for resolving the conflict is to be conscientious and do your best (2012: 211). In the face of disagreement we often don't know the best way to resolve the conflict and we don't always know what the most conscientious response is. In this way, her discussion may strike some as insensitive to the nuances surrounding the problem.

CONCLUSION

Although I agree with several of Zagzebski's main conclusions, the reasoning she uses to arrive at these conclusions I find wanting in the ways I've mentioned here. By taking the egoist as her main opponent, Zagzebski ends up treating several complex issues with too blunt an instrument. Arguably, the notion of conscientiousness does too much work for her account given the insufficient discussion of what exactly it is to be conscientious. Nevertheless, she points us in the direction of topics worthy of discussion and has constructed a broad framework within which to grapple with these questions. Any reader will find something of interest to mull over in this book, and it will easily become required reading for anyone working on the concept of epistemic authority.

EPISTEMIC AUTHORITY, PRE-EMPTION, AND NORMATIVE POWER

BENJAMIN MCMYLER

Texas A&M University

I. INTRODUCTION[1]

Linda Zagzebski's *Epistemic Authority* is an expansive book, exploring a range of issues from the nature of the self, agency, and trust, to the problems of epistemic circularity and peer disagreement. The heart of the book, however, that which ties all of these disparate topics together, is an exploration and defence of the idea that there is such a thing as genuinely epistemic authority, a kind of authority over belief that genuinely parallels practical authority, or authority over action. Insofar as the topic of authority is a central concern of social and political philosophers, it is surprising that this topic has received relatively little attention from social epistemologists. Zagzebski's book goes a long way towards rectifying this situation, and as such it will undoubtedly serve as a touchstone for future work.

Zagzebski's account of epistemic authority is formulated so as to parallel what can safely be called the standard contemporary account of practical authority, Joseph Raz's service conception of authority.[2] According to Raz, practical authorities perform the service of mediating between agents and the reasons for action that apply to them. Authorities perform this service by issuing directives, such as laws, orders, or commands, that provide agents with reasons for action that are supposed to both reflect and replace the reasons for action that apply to these agents independent of the directive. Raz calls such reasons for action 'pre-emptive' reasons. The notion of pre-emption is supposed to explain the intuitive respect in which practical authorities purport to be in a position to settle for us the

[1] I am indebted to Linda Zagzebski for many discussions of this material, as well as to the comments of an anonymous referee.

[2] The canonical statement of this position is Raz (1986).

question what to do. For Raz, authorities settle practical questions for us by giving us pre-emptive reasons for action.

Following Raz, Zagzebski states that what is essential to authority is the normative power to generate pre-emptive reasons, and she holds that such normative powers can be powers to generate pre-emptive reasons *for belief* just as much as for action.

> What is essential to authority is that it is a normative power that generates reasons for others to do or believe something preemptively ... A preemptive reason is a reason that replaces other reasons the subject has. Believing what another person believes or tells me preemptively is parallel to doing what he tells me to do preemptively. In both cases what the authority does gives me reason to believe or do something that replaces my other reasons relevant to the belief or act. (Zagzebski 2012: 102)[3]

Raz himself seems to accept that there can be epistemic authority that meets the general conditions of his service conception. In a recent paper he writes:

> Just as with any practical authority, the point of theoretical authority is to enable me to conform to reason, this time reason for belief, better than I would otherwise be able to do. This requires taking the expert advice and allowing it to pre-empt my own assessment of the evidence. If I do not do that, I do not benefit from it. (Raz 2009: 155)

Raz here claims that 'expert advice' can provide pre-emptive reasons for belief. I take it that what Raz calls expert advice is what epistemologists would call expert testimony, for example, a climate scientist's telling me that global warming is occurring and is largely the result of human activity. Raz is claiming that such testimony can provide pre-emptive reasons for belief and that agents like the climate scientist thus amount to epistemic (or theoretical) authorities.

Interestingly, however, Zagzebski's initial cases of epistemic authority are not cases of expert or authoritative testimony. Zagzebski's initial defence of epistemic authority in Chapter 5 of her book concerns the authority of belief, not testimony. It concerns cases in which I trust the way in which someone else gets her belief in a particular domain more that I trust the way in which I would get the belief myself. As she puts it,

[3] In a similar vein, Lawlor (2013) argues that the speech act of assurance purports to provide hearers with exclusionary or pre-emptive reasons for belief. I argue for a similar parallel between testimony and authoritative practical directives in McMyler (2011).

'In cases of these kinds the conscientious thing to do is to let the other person stand in for me in my attempt to get the truth in that domain and to adopt his belief. This in broad outline is what I mean by epistemic authority.' (2012: 105) Zagzebski notes that this is not the most natural parallel to the case of practical authority. She claims that the authority of testimony, the topic of Chapter 6, is a stronger form of authority that more closely resembles authority over action (2012: 119). Nevertheless, she thinks that the authority of belief bears sufficient resemblance to practical authority for it to count as a genuine form of authority. Zagzebski thus proposes that there are two general types of epistemic authority, the authority of belief and the authority of testimony, both of which deserve to be characterized as forms of authority in virtue of their being normative powers to generate for others pre-emptive reasons for belief.

I think that Zagzebski is right that there is such a thing as epistemic authority, where this is understood as a kind of authority over belief that robustly parallels authority over action. With this much agreement in mind, however, I want to raise three worries concerning the details of Zagzebski's account, all of which bear on her central contention that epistemic authority is a normative power to generate pre-emptive reasons for belief. First, it is difficult to see how Zagzebski's account of the authority of belief meets the conditions that Raz proposes on the possession of a normative power. If authority is a normative power over others, and if we accept Raz's conception of the nature of normative powers, then it is difficult to see how there can be such a thing as the authority of belief. This is not a problem for Zagzebski's account of the authority of testimony. However, second, if we accept Raz's conception of the nature of pre-emptive reasons, it is difficult to see how there can be such a thing as a pre-emptive reason for belief. Raz holds that pre-emptive reasons are a species of what he calls 'second-order reasons', but given the nature of second-order reasons, I don't see how there can be reasons that are epistemically second-order. This is a problem for both Zagzebski's account of the authority of testimony and her account of the authority of belief. Finally, third, even if we can make sense of the notion of pre-emptive epistemic reasons, it isn't clear that the normative power to give pre-emptive reasons is actually sufficient for authority. This is a problem for the Razian framework for understanding authority in general, be it epistemic or practical. I end on a positive note, suggesting that Zagzebski's discussion of the authority of testimony points to

considerations that take us beyond the official Razian framework for understanding authority, considerations that might be employed to address the second and third problems outlined here.

II. BELIEF AND NORMATIVE POWER

As I have said, Zagzebski's initial defence of epistemic authority in Chapter 5 concerns the authority of belief. Generally speaking, another's belief is authoritative for me when I conscientiously trust the way in which the other gets her belief more than I would the way in which I would get my own belief were I to attempt to determine what to believe myself. In such cases, Zagzebski claims that the reasonable thing to do is to let the authority 'stand in for me' in determining what to believe, and I do this by allowing the authority's judgment to pre-empt my own. She formulates this in terms of the following pre-emption thesis for epistemic authority:

> The fact that the authority has a belief p is a reason for me to believe p that replaces my other reasons relevant to believing p and is not simply added to them. (2012: 107)

Zagzebski is aware that this does not directly parallel Raz's account of practical authority, but I think it is worth pausing a moment to see why. Consider in this respect what a parallel pre-emption thesis for practical authority might look like:

> The fact that the authority Φs (or intends to Φ) is a reason for me to Φ (or intend to Φ) that replaces my other reasons relevant to Φ-ing (or intending to Φ) and is not simply added to them.

This is not Raz's pre-emption thesis. Raz's pre-emption thesis for practical authority is instead something like the following:

> The fact that an authority directs (commands, orders, tells) me to Φ is a reason for me to Φ that replaces my other reasons relevant to Φ-ing and is not simply added to them.[4]

[4] Raz writes, 'One thesis I am arguing for claims that authoritative reasons are preemptive: the fact that an authority requires performance of an action is a reason for its performance which is not to be added to all other relevant reasons when assessing what to do, but should exclude and take the place of some of them.' (1986: 46) I have formulated the pre-emption thesis in the text in the first-person in order to parallel Zagzebski's pre-emption thesis. Note that on Raz's formulation, the notion of an authority's 'requiring performance', or as I put it above, directing me to Φ, is integral to the account.

Raz holds that practical authority is a matter of the way in which the *directives* of authorities are capable of providing pre-emptive reasons for action. Zagzebski's pre-emption thesis for epistemic authority does not require the issuing of directives, and as a result, I doubt that it amounts to something that is rightly referred to as a kind of authority.

Note that the practical parallel of Zagzebski's pre-emption thesis does not characterize cases of practical authority. I might very well treat someone that I admire as an exemplar concerning which actions to perform in a given domain without thereby treating her as a practical authority. For example, I might seek to emulate my neighbour's gardening techniques, or the intentions involved therein, without thereby treating my neighbour as having any practical authority over me. In emulating my neighbour, I am not obeying her, and this is because she hasn't told me to do anything. So if the practical parallel of Zagzebski's pre-emption thesis doesn't characterize cases of practical authority, why should we think that the epistemic version characterizes something that deserves to be called authority?

One might suggest that epistemic and practical authority simply differ in this regard. I take it that this is roughly Zagzebski's view. Along these lines, one might suggest that in seeking to emulate my neighbour's gardening techniques I am in fact treating her as an epistemic authority. After all, her actions express her beliefs about what is to be done gardening-wise, and if I conscientiously trust those beliefs more than I would my own beliefs were I to attempt to determine what to do for myself, then I might treat the beliefs that are expressed in her actions as pre-emptive reasons for belief. Zagzebski herself describes a case very much like this as a case of epistemic authority: 'I may want to attend a lecture but might not be sure where the lecture room is. I may then follow a group of people whom I know are going to the lecture. I assume that their behaviour indicates their belief about the location of the room, and when I do so, it is their belief that I take to be authoritative.' (2012: 120)

I think there is reason to resist construing such cases as cases of genuine authority. Recall that Zagzebski accepts Raz's conception of authority as a normative power to generate for others pre-emptive reasons. As she recognizes, having what one says or does treated as a pre-emptive reason is insufficient for possession of such a normative power. Immediately after introducing Raz's account of authority as normative power, she writes:

> I am not suggesting that taking a reason to be preemptive is sufficient for either acting or believing on authority. If you love someone, you might take the fact that he or she asks you to do something as a preemptive reason to do it, but when you do so you are not treating the beloved as an authority. Similarly, it is possible (although less likely), that you will take the fact that the loved one has a certain belief as a preemptive reason to believe it. If so, believing preemptively is not sufficient for believing on authority. We usually do not think that the people we love have a normative power to give us preemptive reasons just because we love them, even if we choose to take their wishes as giving us preemptive reasons. In contrast, authority is such a power. (Zagzebski 2012: 102-103)

Zagzebski here claims that we might rationally or justifiably treat a loved one's request (or belief) as a pre-emptive reason for action (or belief) without this making the loved one a practical (or epistemic) authority. Treating what another does or believes pre-emptively is insufficient for the person's possessing a normative power to generate pre-emptive reasons for others. What more, then, is required for possession of such a normative power?

In *Practical Reason and Norms*, Raz gives the following definition of normative power:

> An act is the exercise of a normative power if, and only if, it is recognized as effecting a normative change because, among other possible justifications, it is an act of a type that, if recognized as effecting a normative change, acts of this type will be generally performed only if the persons concerned want to secure this normative change. (Raz 1999: 103)[5]

There are two things to note about this definition. First, this definition presupposes that the exercise of a normative power is an act, an intentional action. For Raz, to exercise the normative power to give pre-emptive reasons is, paradigmatically, to order, command, or legislate that someone do something. No such actions are present in the case of my neighbour or Zagzebski's pedestrians. My neighbour has not directed me to do or believe anything, and neither have the pedestrians in Zagzebski's example.

But perhaps Raz's definition needs to be modified in order to make room for epistemic authority. Perhaps another person's believing that

[5] See also Raz (1979: 18).

p in a way that is conscientiously recognized by me as being more conscientious than would be my own belief were I to try to form the belief myself is sufficient for possessing the relevant normative epistemic power. We might thus try to modify Raz's definition of normative power by replacing talk of acts and performing actions with talk of beliefs and holding beliefs:

> A belief is the exercise of a normative power if, and only if, it is recognized as effecting a normative change because, among other possible justifications, it is a belief of a type that, if recognized as effecting a normative change, beliefs of this type will generally be held only if the persons concerned want to secure this normative change.

Second, however, not only does Raz's definition of normative power presume that exercises of normative powers are intentional actions, it also holds that these actions must be of a type that are typically performed in order to secure a normative change, in the case of authority, in order to give to or generate for others pre-emptive reasons. Orders and commands are speech acts that are issued with the intention of giving or generating pre-emptive reasons. This is what distinguishes such exercises of normative power from other speech acts, like requests, that might be rationally treated as providing pre-emptive reasons but that are not issued with the intention of doing so. Neither my neighbour's gardening beliefs nor the pedestrian's beliefs concerning the location of the lecture are held in order to generate pre-emptive reasons for others. In fact, it is unclear what it could even mean to hold a belief in order to secure a normative change. The attitude of belief looks like the wrong kind of thing to be the exercise of a normative power, and so believing that p, even though it might be rationally treated as providing pre-emptive reasons, cannot itself be an exercise of authority.

One might object that Raz himself claims that the intention to secure a normative change is not necessary for exercising a normative power.

> Normally only acts done with the intention of producing relevant normative change are recognized as producing it. But this is not always the case, and there are many exceptions particularly in the law or other institutionalized normative systems. One may make a binding contract without realizing that one did, for example. For this reason, the definition [of normative power] turns, not on the intentions with which the act is performed, but rather on the reasons for regarding it as effecting a normative change. (Raz 1999: 104)

If I understand him, what Raz is here claiming is that the exercise of a normative power should be understood not in terms of the intentions of the person or institution exercising the power on the particular occasion but rather in terms of what justifies us in taking what the person or institution does to be an exercise of such a power. On the surface, this sounds congenial to Zagzebski's position. After all, she offers several epistemic parallels to Raz's Normal Justification Thesis showing that epistemic authority can be justified for us in much the same way as practical authority. Raz's Normal Justification Thesis states that 'the normal way to establish that a person has authority over another person involves showing that the alleged subject is likely better to comply with reasons which apply to him (other than the alleged authoritative directives) if he accepts the directives of the alleged authority as authoritatively binding and tries to follow them, rather than by trying to follow the reasons which apply to him directly' (Raz 1986: 53). Zagzebski offers two parallel justification theses for the authority of belief, the second of which states that 'the authority of another person's belief for me is justified by my conscientious judgment that I am more likely to form a belief that survives my conscientious self-reflection if I believe what the authority believes than if I try to figure out what to believe myself' (Zagzebski 2012: 110-111). As I understand it, however, Raz's Normal Justification Thesis simply assumes that we have in view an attempted exercise of a normative power, an attempt to give pre-emptive reasons. The question that the Normal Justification Thesis is meant to answer is the question of what makes such an attempt successful, one that succeeds in giving the kind of reason that it purports to give.[6] It is an account of what makes an authority legitimate. The Normal Justification Thesis does not answer the question of what justifies us in taking something to be an attempted exercise of a normative power. Raz's answer to this question lies in his definition of normative power. We are justified in taking something to be an attempted exercise of a normative power only if it is an act of a type that, if recognized as effecting a normative change, acts of this type will be generally performed only if the persons concerned want to secure this normative change.

In this respect, the reason that one can make a binding contract without realizing it or intending to do so is that one can perform actions, such as signing a document, that are of a type that, if recognized

[6] See, for example, Raz (1986: 53).

as effecting a normative change, acts of this type will be generally performed only if the persons concerned intend to so bind themselves. So while a particular person's intending to secure a normative change isn't necessary for the exercise of a normative power on a particular occasion, for such an exercise to count as such it must be of a type that is recognized as being generally performed only if the persons involved want to secure this change. Beliefs are simply not of this type. So while we might have good reason to treat the beliefs of others as pre-emptive reasons for belief, particularly when Zagzebski's Justification Theses for the Authority of Belief are satisfied, if authority is a normative power to generate pre-emptive reasons for others, it looks like there can be no authority of belief.

III. PRE-EMPTIVE REASONS FOR BELIEF

This problem concerning the notion of normative power does not tell against Zagzebski's account in Chapter 6 of the authority of testimony. In contrast to the case of the authority of belief, we can easily formulate a pre-emption thesis for the authority of testimony that parallels Raz's pre-emption thesis for practical authority:

> The fact that an authority tells me that p is a reason for me to believe that p that replaces my other reasons relevant to believing that p and is not simply added to them.

Just as telling an audience to Φ is a speech act that purports to give the audience a pre-emptive reason to Φ, so telling an audience that p can be construed as a speech act that purports to give the audience a pre-emptive reason to believe that p. If the case of my neighbour is modified such that it involves her telling me that such and such is the thing to do in my garden, and if Zagzebski's case of the pedestrians is modified such that it involves their telling me the location of the lecture, then it seems much more plausible that we have in view cases in which I might believe these things on the authority of the relevant speakers. In telling me that such and such is the case, the relevant speakers can be construed as performing actions that satisfy Raz's conditions on the exercise of a normative power, actions of a type that, if recognized as effecting a normative change, acts of this type will be generally performed only if the persons concerned want to secure this normative change. And as we

have seen, Raz himself seems to accept that such authoritative testimony can meet the general conditions of his service conception of authority.

Still, I would like to raise a further problem for this account, one concerning the notion of pre-emption itself. Raz provides a fairly detailed account of the nature of pre-emptive reasons, and I worry that these details make it difficult to see how there can be such a thing as a pre-emptive epistemic reason.

For Raz, pre-emptive reasons are a species of what he calls *second-order reasons*. The point of introducing this category of second-order reasons is to illuminate differences in the kinds of conflict that can arise between reasons. While first-order reasons can defeat and outweigh one another, Raz holds that second-order reasons interact with other reasons in a quite different way. Second-order reasons are reasons to act for first-order reasons or to refrain from acting for first-order reasons (1999: 39). Second-order reasons do not affect the strength of first-order reasons. Rather, they determine whether certain first-order reasons are to be acted upon, and as such they cannot be simply added to the balance of first-order reasons. This is what it means to say that second-order reasons *replace* other reasons. Raz calls negative second-order reasons, reasons for refraining from acting for certain first-order reasons, *exclusionary* or *pre-emptive* reasons. Pre-emptive reasons serve to exclude or pre-empt other reasons by being reasons for not acting for these other reasons.

In addition to their pre-emptive nature, the reasons for action provided by authoritative directives are themselves first-order reasons for performing the root action. An authority's ordering me to Φ is for me both a first-order reason to Φ and a second-order reason not to act for certain other conflicting reasons. In this respect, the pre-emptive nature of an authoritative directive serves to 'protect' the first-order reason provided by the directive itself. Raz calls reasons that are both first-order reasons and second-order exclusionary or pre-emptive reasons *protected reasons* (1979: 18).

Second-order reasons are not simply reasons for performing or refraining from performing the root action. If they were, then they wouldn't be adequately distinguished from first-order reasons. What makes second-order reasons second-order is that they are reasons for performing a kind of higher-order action, the action of acting for a reason or intentionally refraining from acting for a reason.

> There is one important point to bear in mind concerning second-order reasons: They are reasons for action, the actions concerned being acting

> for a reason and not acting for a reason. If P is a reason to Φ then acting for the reason that P is Φ-ing for the reason that P. Not acting for P is not Φ-ing for the reason that P. (Raz 1979: 17)

If I understand him correctly, Raz takes Φ-ing for the reason that P and intentionally refraining from Φ-ing for the reason that P to be themselves actions, call them acts of Ψ-ing. We might thus call Ψ-ing *a second-order action*. Ψ-ing is the second-order action of Φ-ing for a reason or refraining from Φ-ing for a reason. Pre-emptive reasons are thus reasons for Ψ-ing, where to Ψ is to refrain from Φ-ing for certain conflicting first-order reasons.

By construing pre-emptive reasons in this way, Raz is in a position to claim, quite plausibly, that authoritative directives do not pre-empt subjects from considering or deliberating about the conflicting first-order reasons that the directive is meant to exclude. Subjects can deliberate all they want about such conflicting reasons. Authoritative directives are only meant to pre-empt subjects' acting on the excluded reasons.[7]

The fact that pre-emptive reasons are reasons for what I am calling second-order actions puts pressure on the idea that there can be pre-emptive reasons for belief.[8] If there are pre-emptive reasons for belief, then presumably they are not second-order reasons for action. Instead, they must be second-order reasons for some kind of doxastic parallel to second-order actions. But is there such a parallel? Second-order actions are acts of acting or refraining from acting for reasons, but is there anything like a higher-order doxastic state that is itself an instance of believing for a reason or intentionally refraining from believing for a reason? I cannot see that there is. Of course, there are second-order beliefs in the sense of beliefs that have other beliefs as their content, but this isn't what we need. What we need is something that parallels a second-order action, an act of Ψ-ing, where to Ψ is to Φ for a reason or intentionally refrain from Φ-ing for a reason. Believing that p for a reason or refraining from believing that p for a reason is not itself an instance of a higher-order doxastic state. I thus don't see how there can be such a thing as a doxastic parallel to second-order actions.

---

[7] See especially the discussion in Raz (1989).

[8] For doubts about the notion of pre-emptive or exclusionary reasons for action see Gans (1986). While I am here assuming that we can make sense of the notion of second-order reasons for action, I suspect that some of the considerations adduced here might be developed in such a way as to call into question the general category of second-order reasons.

Here is another way of making the same point. Consider what a reason for believing for a reason or refraining from believing for a reason might be. Here it will be helpful to introduce some further non-Razian terminology.[9] Let's take *a reason* to be a consideration that bears, not on an action or attitude, but on a question. A consideration that bears on the question whether to Φ, where Φ-ing is an action, is a practical reason, and a consideration that bears on the question whether p, where p is a proposition, is an epistemic reason. Second-order reasons for action can thus be construed as considerations that bear on the question whether to act for a reason. Insofar as Raz construes acting for a reason as itself a kind of action, what I've called a second-order act of Ψ-ing, the question whether to act for a reason, whether to Ψ, can be construed as a question of the form whether to Φ. But what are we to make of second-order reasons for belief? Second-order reasons for belief would be considerations that bear on the question whether to believe that p for a reason. But what is the form of the question *whether to believe that p for a reason*? To believe that p for a reason is not a proposition. The question whether to believe that p for a reason thus is not a question of the form whether p, and since it is not a question of the form whether p, considerations that bear on this question cannot be epistemic reasons.

The concept of second-order reasons for belief thus looks to me to be incoherent. If there are such reasons, then presumably they are epistemic reasons, considerations that bear on a question of the form whether p.[10] But I cannot see how there can be, as it were, second-order questions of this form. Second-order reasons for action are considerations that bear on the question whether to Ψ, where to Ψ is to Φ for a reason. Second-order epistemic reasons would be considerations that bear on a question

[9] I here draw on the conception of reasons and rational agency recently defended by Pamela Hieronymi. See especially Hieronymi (2005) and (2006).

[10] Perhaps second-order reasons for belief can be construed as practical reasons, as what are sometimes called 'pragmatic' reasons for belief. They would then be reasons for performing actions (including mental actions) designed to bring about believing for a reason or not believing for a reason. For example, a speaker's testimony that p might be construed as both a first-order epistemic reason for believing that p and a practical reason for acting in ways designed to, say, prevent one from failing to believe that p for certain other conflicting reasons. On such a construal, however, I don't think we have something that is aptly construed as a second-order reason. Instead, we have a consideration that is both a first-order epistemic reason and a first-order practical reason. This amounts to a serious departure from the official Razian account of authority. Moreover, while I cannot consider such a proposal in any detail here, I doubt that it will yield a plausible conception of the nature of epistemic authority.

of the form whether p, where p was itself something like believing for a reason. But believing for a reason isn't a proposition. I therefore tentatively conclude that, insofar as pre-emptive reasons for belief are supposed to be second-order epistemic reasons, there can be no pre-emptive reasons for belief.

Raz has a quite detailed conception of the nature of pre-emptive reasons. I have argued, *pace* Raz himself, that these details do not leave room for pre-emptive reasons for belief. Notably, however, Zagzebski does not discuss many of the details of the Razian conception of pre-emption, including the notion of second-order reasons that I have relied upon heavily here, so perhaps she has a slightly different conception of pre-emption in mind. I have said nothing about a non-Razian distinction that Zagzebski introduces between first-person or deliberative reasons and third-person or theoretical reasons, and there is at least one point at which Zagzebski suggests that this distinction is helpful for understanding the way in which pre-emption works (2012: 114). Perhaps the distinction between deliberative and theoretical reasons can be employed so as to develop an alternative account of pre-emption that will make room for pre-emptive reasons for belief. At this point, however, I do not see how exactly this would go.

IV. AUTHORITIES AND INSTRUMENTS

Even if we can make sense of the notion of pre-emptive reasons for belief, however, I have a final worry. I admit that this worry is a bit amorphous, but it seems to me important. The philosophical problem of authority is often construed as the problem of how it can ever be morally or rationally justified to submit oneself to the will of another in the way that is characteristic of deference to authority. The general Razian strategy for addressing this problem, I take it, is to argue that, as long as certain conditions are met, conforming to the directives of authorities is an indirect way of conforming to the reasons for action that exist anyway. As Raz has recently put it:

> In postulating that authorities are legitimate only if their directives enable their subjects to better conform to reason, we see authority for what it is: not a denial of people's capacity for rational action, but simply one device, one method, through the use of which people can achieve the goal (*telos*) of their capacity for rational action, albeit not through its direct use. (2009: 140)

Similarly, Zagzebski argues that, while there might be occasions on which we value deciding for ourselves more than we value truth, believing on authority is an indirect means of attaining the epistemic ends of attaining truth and avoiding falsity. To the extent that we value getting the truth, we are often better off relying on epistemic authorities than we are trying to determine what to believe for ourselves. We thus have a straightforward defence of deference to epistemic authority as simply the most rational or epistemically conscientious thing to do.

While I think that this is all very plausible, I worry that it sidesteps the problem of authority rather than squarely addressing it. For many philosophers, the philosophical problem of authority is not simply a problem concerning whether deference to authority can be instrumentally justified. Instead, the problem seems to concern whether the notion of deference to epistemic or practical authority is even coherent. Just as an example, consider the way in which Herbert Marcuse frames the issue in the introduction to his *A Study on Authority*:

> The authority relationship, as understood in these analyses, assumes two essential elements in the mental attitude of he who is subject to authority: a certain measure of freedom (voluntariness: recognition and affirmation of the bearer of authority, which is not based purely on coercion) and conversely, submission, the tying of will (indeed of thought and reason) to the authoritative will of an Other. Thus in the authority relationship freedom and unfreedom, autonomy and heteronomy, are yoked in the same concept and united in the single person of he who is subject. (2008: 7)

Marcuse here claims that there is a kind of paradox involved in the very concept of obedience or deference to authority. Authorities aim at a response from their subjects that is somehow simultaneously free and unfree, that involves freely willing to not will for oneself or freely judging to not judge for oneself. This sense of paradox in the notion of subjecting or submitting oneself to authority is something that gets washed out in both Raz and Zagzebski's accounts. Perhaps it is appropriately washed out. Perhaps there is nothing here but confusion. But it seems to me that there is something correct and important about what Marcuse, like many others, is gesturing at.[11] The question is how to understand it.

[11] Similar sentiments are expressed by philosophical anarchists like Godwin (1971) and Wolff (1970) and by defenders of authority like Arendt (1954).

On the general Razian view, I take it that the notion of pre-emption is supposed to capture whatever sense there is in the idea that deference to authority involves subjecting or submitting oneself to another.[12] On Zagzebski's epistemic version of this view, epistemic authorities provide reasons that replace my other reasons relevant to believing. In believing that p on the authority of someone else, I am subjecting myself to the judgment of the other in that I am believing that p pre-emptively. The problem, however, is that on both Raz and Zagzebski's accounts, it can look like pre-emption comes rather easily. A reason is pre-emptive just so long as it is one that excludes a belief or action being based on certain other conflicting reasons. In this respect, lots of considerations might amount to pre-emptive reasons. For Raz, not only authoritative directives but other kinds of mandatory rules, promises, and even decisions provide pre-emptive reasons. We've seen that one might be justified in treating a speaker's request or advice as a pre-emptive reason for action, and it seems that we might also be justified in treating things like the readings of ordinary instruments pre-emptively. At the end of Chapter 5, Zagzebski writes that her account of the authority of belief 'applies to inanimate objects like GPS systems, thermometers, and other instruments' (2012: 119). A thermometer might thus provide pre-emptive reasons for belief concerning the temperature in the room, and a car's GPS device might provide pre-emptive practical reasons for taking the next exit.

I doubt that Raz would be willing to count such instruments as authorities. If authority is construed as a normative power to give or generate for others pre-emptive reasons, and if normative powers are defined in roughly the way that Raz defines them, as requiring the performance of an intentional action, then it will look like only agents can be genuine authorities.[13] Still, it isn't clear from Raz's account why exactly agency matters here. His defence of authority focuses on the way in which pre-emption is a means of maximizing conformity with reason, and this defence doesn't appeal to anything explicitly agential. The considerations that he claims provide pre-emptive reasons are intentional actions, but it isn't clear why they need be. In a way, this contributes to the strength of the Razian defence of authority. Authority looks no

[12] See, for example, Raz's claim that the philosophical anarchist simply overlooks the possibility of the existence of second-order reasons (1979: 27).

[13] Perhaps instruments can be construed as having a kind of authority that is derived from the authority of their designers, but then it looks like we need a distinction between original and derived authority.

more morally or rationally problematic than any other indirect means of securing conformity with reason, but at the same time, I worry that we've lost touch with what was supposed to be so distinctive, and distinctively worrying, about authority. The paradoxical nature of authority pointed to by writers like Marcuse seems fundamentally to do with the notion of freely submitting oneself to another agent.

Consider in this respect what distinguishes our reliance on instruments from our reliance on practical and epistemic authorities. In the case of instruments that provide reasons for action, like, for example, a car's GPS device, we are often justified in treating what the device indicates as a reason for action. We might even be justified in treating what the device indicates as a pre-emptive reason, as a reason that is not simply added to the balance of reasons but that actually excludes certain other conflicting reasons. Even if we treat what the device indicates pre-emptively, however, we are not obeying the device (though we might sometimes choose to talk in this way). We are not treating the device as we would another person whom we recognize as being in a position to settle for us the question what to do. We are not subjecting our wills to the device, and it is this subjection to others that renders authority both distinctive and distinctively problematic.

This suggests that the distinctiveness of authority cannot be explained in terms of pre-emption alone, in the way that Raz tends to explain it. If ordinary instruments can provide pre-emptive reasons, as Zagzebski suggests, and if in treating them pre-emptively we are not subjecting our wills to the instrument in the way that obedience to authority involves subjecting our will to the authority, then the distinctiveness of authoritative directives cannot consist in their pre-emptive nature alone.

This is a general worry about the sufficiency of the Razian account of authority, but it has a more specific application in the case of epistemic authority. I have suggested that the difference between the indications of a practical instrument like a car's GPS device and the directives of legitimate practical authorities is that the latter purport to settle practical questions for us, thereby subjecting us to their wills, in a way that the former do not. The notion of authorities settling questions for us is, of course, in need of further explanation, but I have suggested that pre-emption alone might be insufficient to explain it. If we turn to the case of epistemic authority, however, one might wonder whether this notion even applies. There is, I think, a pretty clear intuitive distinction between

following the directions of a GPS device and obeying the directives of a legitimate practical authority, but one might wonder whether there is such a clear distinction between, for example, believing the readings of a thermometer and believing the testimony of a climate scientist. We might treat both pre-emptively, but does believing the testimony of the climate scientist really involve submitting ourselves to the scientist in a way that believing the readings of a thermometer does not? If not, then it isn't clear that testimony involves a kind of authority that robustly parallels practical authority.

Zagzebski seems to think that there *is* a genuine sense in which testimony, like authoritative practical directives, purports to settle theoretical questions for us in a way that involves our submitting ourselves to the authority. At the beginning of Chapter 6, she argues that the authority of testimony brings with it an interpersonal dimension that is lacking in the case of the authority of belief. The speech act of telling an audience that p differs from other speech acts that aim to influence the beliefs of others in that telling involves an assumption of responsibility on the part of the speaker for the conscientiousness of the audience's belief.

> Telling is a two-way street. The teller asks for trust and counts on the recipient to trust her. In return, she assumes the responsibility that goes with that trust, taking upon herself the epistemic burden of believing in a conscientious fashion, and doing so not only for herself, but for the recipient. (Zagzebski 2012: 124)

In telling an audience that p, a speaker aims for the audience to believe her that p, to believe that p on the speaker's authority, where this involves the audience's ceding to the speaker at least partial responsibility for the conscientiousness of her belief. Perhaps something similar is true of practical authority. In telling an audience to Φ, a speaker aims for the audience to obey her, where this involves the audience's ceding to the speaker at least partial responsibility for the appropriateness of her action, for her Φ-ing being the thing to do in the situation. In this way, believing that p on the authority of a speaker might be construed as allowing the speaker to settle for one the question whether p in a way that parallels that in which obeying a speaker's command involves allowing the speaker to settle for one the question what to do. The notion of an authority's settling questions for others and thereby subjecting others to her will or judgment is here cashed out in terms of the responsibility

that the authority assumes, and the subject cedes to the authority, for the conscientiousness of the audience's belief or action.

Such an account of the parallel between theoretical and practical authority has several virtues. First, it succeeds, I think, in distinguishing obedience or deference to authority from reliance on ordinary instruments. Responsibility for belief and action can only be distributed between beings capable of such, between agents. Even if we treat the readings of an instrument pre-emptively, we are not thereby ceding responsibility for our belief or action to the instrument. Second, and relatedly, this account appears to be in a better position than the general Razian position to make sense of the paradox of authority depicted in the passage from Marcuse. On the general Razian view, pre-emption is an indirect means of maximizing conformity to reason. On the alternative outlined here, however, authority has a distinctive impact on the agency of others. Authorities purport to take over for others the activity of settling theoretical and practical questions, and while this might very well have the effect of maximizing others' conformity to reasons, one might worry that taking over aspects of others' agential activity remains an affront to their status as rational agents. Third, this account of the parallel between epistemic and practical authority avoids the problem that I have raised concerning the application of the notion of pre-emption to the realm of belief. As far as I can see, there is nothing standing in the way of thinking that responsibility for belief (assuming, I think plausibly, that there is such a thing) can be distributed between authority and subject in a way that parallels that in which responsibility for action can be so distributed. As such, this account makes room for a robust parallel between epistemic and practical authority.

Zagzebski's claims concerning the interpersonal dimension of the authority of testimony do not help to solve the problem that I have raised concerning the authority of belief. It is hard to see how either the case of my neighbour or Zagzebski's case of the pedestrians involve the interpersonal ceding and accepting of responsibility involved in an authority's settling for another a theoretical question. Nevertheless, I think that they do go some distance towards solving the other two problems that I have raised. They do so, however, only by going beyond the official Razian framework for understanding authority. They are consistent with the idea that authority is, at least in part, a normative power to give pre-emptive reasons, but the appeal to pre-emption doesn't itself explain these interpersonal features. If this is right, then

the interpersonal dimension of authority might be something that needs to feature more centrally in our understanding of both epistemic and practical authority.

BIBLIOGRAPHY

Arendt, Hannah. 1954. 'What is Authority?' in *Between Past and Future* (New York: The Viking Press)

Gans, Chaim. 1986. 'Mandatory Norms and Exclusionary Reasons', *Philosophy*, 15: 373-394

Godwin, William. 1971. *Enquiry Concerning Political Justice* (Oxford: Clarendon Press)

Hieronymi, Pamela. 2005. 'The Wrong Kind of Reason', *The Journal of Philosophy*, 102: 437-457

Hieronymi, Pamela. 2006. 'Controlling Attitudes', *Pacific Philosophical Quarterly*, 87: 45-74

Lawlor, Krista. 2013. *Assurance: An Austinian View of Knowledge and Knowledge Claims* (Oxford: Oxford University Press)

Marcuse, Herbert. 1972. *A Study on Authority* (New York: Verso)

McMyler, Benjamin. 2011. *Testimony, Trust, and Authority* (New York: Oxford University Press)

Raz, Joseph. 2009. 'The Problem of Authority: Revisiting the Service Conception', in *Between Authority and Interpretation* (New York: Oxford University Press)

Raz, Joseph. 1999. *Practical Reason and Norms* (New York: Oxford University Press)

Raz, Joseph. 1989. 'Facing Up: A Reply', *Southern California Law Review*, 62: 1153-1235

Raz, Joseph. 1986. *The Morality of Freedom* (New York: Oxford University Press)

Raz, Joseph. 1979. *The Authority of Law* (New York: Oxford University Press)

Wolff, Robert Paul. 1970. *In Defense of Anarchism* (Berkeley: University of California Press)

Zagzebski, Linda Trinkaus. 2012. *Epistemic Authority: A Theory of Trust, Authority, and Autonomy in Belief* (New York: Oxford University Press)

TRADITION AS TRANSMISSION: A PARTIAL DEFENCE

JOHN SCHWENKLER

Florida State University

DISCLAIMER

Linda Zagzebski's *Epistemic Authority*[1] is a wonderful book, and I learned a great deal from reading it. But philosophers are trained to disagree, so disagree I shall. (I would be glad to learn that the distance between us is more apparent than real.)

I. TRADITION AS TRANSMISSION

In discussing religious authority in ch. 9 of her book, Zagzebski distinguishes three conceptions of how divine revelation is transmitted through a religious tradition. According to one of these conceptions, which henceforth I'll call *Tradition as Transmission*, a religious tradition consists solely in chains of testimony that stretch back to an original encounter with some past events.[2] Zagzebski raises two objections to this model, the first of which is that it fails to explain how such a tradition can be a source of knowledge:

> On the chain model it is crucial that the chain is unbroken and that the transmission is accurate. This model assumes that *what* is transmitted

[1] Linda Trinkaus Zagzebski, *Epistemic Authority: A Theory of Trust, Authority, and Autonomy in Belief* (Oxford: Oxford University Press, 2012). References to this book will be bracketed in the text.

[2] As Zagzebski puts it, on this model 'the transmission of a tradition is reducible to chains of testimony. What justifies belief in what the tradition transmits is a relation to something that happened at the origin – for example, the experience of Moses on Sinai, the Apostles' experience of Jesus Christ, or the revelation of Muhammad, and what happened at the origin is understood as immediate contact with the divine, the experience of which is transmitted by oral and written testimony to the present' (p. 193).

> remains the same as what it was at the point of origin. Revelation in this model is fragile because every time it passes from hand to hand, it runs the risk that some of it gets lost or distorted. On this model nobody can be as justified in a belief acquired through the mechanism of the tradition than the person who had divine contact at the beginning of the chain. The nearer one is to the source of revelation, the more complete and accurate the knowledge. Given that we are so far in time from the origin of the chain, the most we can do is to study old sources in greater depth, or perhaps discover ancient books that were lost at some point along the chain. (p. 193)

As she notes, Zagzebski's argument here parallels a passage in Locke's *Essay*, where he writes that 'any Testimony, the farther off it is from the original Truth, the less force and proof it has':

> The Being and Existence of the thing itself, is what I call the original Truth. A credible Man vouching his Knowledge of it, is a good proof: But if another equally credible, do witness it from his Report, the Testimony is weaker; and a third that attests the Hear-say of an Hear-say, is yet less considerable. So that *in traditional Truths, each remove weakens the force of the proof*: And the more hands the Tradition has successively passed through, the less strength and evidence does it receive from them. (Bk. IV, ch. xvi, §10)[3]

For Zagzebski as for Locke, the fallibility of human testimony means that a person who believes a proposition on the say-so of another is necessarily in a worse epistemic position than the person whose testimony she believes. In Zagzebski's view, this shows that an adequate model of religious tradition must take it to involve more than mere chains of testimony, at least if it is to explain the possibility of genuine religious knowledge.

As I read him, Thomas Aquinas endorses the opposite position in his discussion of sacred doctrine (what we today call 'theology') in the first chapter of the *Summa Theologiae:*

[3] For a similar position with respect to demonstrative reasoning, see Book I, Part IV, Section 1 of Hume's *Treatise* ('Of scepticism with regard to reason'). Hume's argument there suggests a 'bad company' objection to Locke's and Zagzebski's: inference is fallible, so if my belief that *p* is based on inference, then by parity of reasoning I should know it less well than I know what it is based on. Some philosophers would accept this consequence, but I expect that Zagzebski will not. For related discussion, see Tyler Burge, 'Content Preservation', *The Philosophical Review*, 102 (4) (1993), 457-488.

> Sacred doctrine is a science. Yet bear in mind sciences differ from each other. Some work from first principles known by the natural light of the intellect – such as arithmetic, geometry, and the like. Others, however, work from principles known by the light of a higher science. Optics, for instance, begins from geometrical principles, and music proceeds from arithmetical ones.
>
> Sacred doctrine is a science in the second sense here, for it proceeds from principles made known by a higher science – that of God and the blessed. So, just as music relies on principles taken from arithmetic, sacred doctrine relies on principles revealed by God. (*ST* I, q. 1, a. 2, c.)[4]

In calling theology a science (*scientia*), Aquinas means to distinguish it from uncertain or merely probable bodies of knowledge or opinion, categorizing it instead as the sort of demonstrative understanding described by Aristotle in the *Posterior Analytics*.[5] This generates a puzzle, however, since according to Aristotle such understanding must proceed from self-evident first principles, and even sensory perception is ruled out as a source of understanding in this strict sense. Aquinas' response to the puzzle is contained in the passage quoted above: he holds that one body of scientific knowledge will sometimes 'borrow' some of its first principles from another *scientia*, as the principles of music are not proved within music itself, but rather within mathematics. Applied to the present case, Aquinas' claim is that the *scientia* of divine things that we have through sacred theology is based on God's immediate knowledge of himself, which is shared with human beings through God's special revelation and the teaching of the church. Yet he insists that this does not render theology any more 'fragile' than other bodies of knowledge, but rather that it is made *more* certain through this mediation than it would be if it had been based on human reason alone:

> We reckon one theoretical science to be more noble than another first because of the certitude it brings ... The science of sacred doctrine surpasses the others [on this count], because theirs comes from the natural light of human reason, which can make mistakes, whereas sacred doctrine is held in the light of God's knowledge which cannot be mistaken. (*ST* I, q. 1, a. 5, c.; and cf. II-II, q. 2, a. 4, c.)

---

[4] In Aquinas, *Summa Theologiae: Questions on God*, ed. Brian Davies and Brian Leftow (New York: Cambridge University Press, 2006).

[5] See Geoffrey Turner, 'St Thomas Aquinas on the 'Scientific' Nature of Theology', *New Blackfriars*, 78 (2007), 464-476.

While Aquinas accepts that the knowledge of God that we attain in this way is less perfect than knowledge gained through epistemically unmediated contact with the divine (see e.g. *ST* I, q. 12, a. 11), the explanation of this has to do with the inability of humans in our present state to understand divine things except by comparing them to created ones, and not with any lack of certainty arising from the mediation of chains of testimony. Unlike Zagzebski, Aquinas sees no problem in the idea that such chains are able to transmit religious knowledge.

I believe that Aquinas' view of this matter is correct, and that the model of religious tradition in terms of testimonial transmission is perfectly able to explain how later generations in a religious tradition can have knowledge at least as secure as that of their ancestors. Section II will argue for this, inspired by some arguments in Tony Coady's seminal work *Testimony: A Philosophical Study*.[6] In the concluding section, I will consider Zagzebski's other objection to Tradition as Transmission, arguing that it is more successful than this first one.

II. BEYOND TELEPHONE

As I see it, Zagzebski's initial argument against the model of Tradition as Transmission requires construing all chains of testimony as similar in structure to a familiar children's game:[7]

> **Telephone:** People are arranged in a line. Someone whispers a message to the first person in line, who whispers it to the second, and so on down. Each person's whisper is inaudible to everyone but the person she is whispering to. When the message reaches the end of the line, it's reported to the whole group.

In *Telephone*, the silliness of the context and the twin difficulties of whispering clearly and making out what is being whispered to you interact to make it unlikely that the message will pass through whole and undistorted. As Zagzebski and Locke both note, these risks arise at

[6] C. A. J. Coady, *Testimony: A Philosophical Study* (Oxford: Clarendon Press, 1992), ch. 11. It's worth noting that Coady's brief discussion of Aquinas in pp. 16-17 of *Testimony* gets his view flatly wrong, and reads Aquinas as holding that faith cannot be a source of knowledge. (Part of the fault lies with the translation that Coady is working from.) A more detailed discussion of how Aquinas' views relate to contemporary work on testimony will have to wait for another essay.

[7] As Coady writes, the Lockean conception of hearsay 'assimilates transmission to mere mimicry, like a series of parrots imitating each other' (*Testimony*, p. 221).

each link in the chain, and thus the longer the chain is, the less reliable it will be.

But real-life testimony, including in the context of religious traditions, is usually quite unlike *Telephone*, which after all is just a game designed to result in a silly outcome. Instead, in everyday life our practice of testimony borrows features from the following possible variations of the game:

> **Reliable:** Like *Telephone*, but everyone in the line has been selected because they are very good at whispering clearly and discerning what is whispered to them, and have no inclination to mess things up on purpose.
>
> **High-Stakes:** Like *Telephone*, but not just for fun: the message is seen as very important, and each person in line has an incentive not to get things wrong. This leads them to listen very carefully, whisper as clearly as they can, and never distort the message on purpose.

The elements of *Reliable* and *High-Stakes* that are missing from *Telephone* make it much less likely that any particular act of transmission will distort the original message, and so increase the reliability of the transmitted signal. And it seems clear that the transmissional practices of most religious traditions have features that mirror each of these: in general, the only individuals licensed to speak authoritatively about doctrinal matters are those with some kind of specialized training, and the matters under discussion are serious enough to the participants in the practice that there should be a strong incentive to transmit them accurately.[8]

But that is not all. Consider now the following further variations on the original:

> **Criss-Crossing:** Like *Telephone*, except that the chain of transmission isn't simply linear: instead of *A* relaying the message to *B*, *B* to *C*, and so on, there will be cases where a member of the chain whispers the message to more than one person, or receives the message from more than one source.[9]
>
> **Convergence:** Like *Telephone*, except that the initial message is whispered to several different people, each of whom begins a chain that converges on a single person at the end.[10]

[8] Coady makes a similar point in *Testimony*, p. 216. Of course things may go the other way, too: recognizing the vital importance of a religious tradition might tempt those who guard it to distort its content in various ways, perhaps to serve their own ends.

[9] For a similar suggestion, see Coady, *Testimony*, pp. 214-215.

[10] Compare Coady, *Testimony*, pp. 212-213.

Back-Tracking: Like *Telephone*, except that sometimes the message passes back through someone earlier in the line, who has the opportunity to correct the message if it has been distorted.

Once again, the features of *Criss-Crossing*, *Convergence*, and *Back-Tracking* that make them different from *Telephone* also make the messages they transmit much less fragile. Specifically, in *Criss-Crossing* and *Convergence* there is the possibility of corroboration or non-corroboration[11] that brings to light potential errors or enables participants to be more confident in the message they are passing along, and the additional mechanism in *Back-Tracking* constitutes a straightforward way to correct errors. And as before, the transmissional practices of religious traditions involve many of these elements, in contrast to the purely linear model suggested by *Telephone*.

Here is one further set of variations that do even more to increase reliability:

Double-Checking: Like *Telephone*, except that each member of the chain is permitted to overhear what the next one says, and to correct her if the message has been transmitted wrongly.

Conferral: Like *Criss-Crossing*, but when multiple people hear the same message they are permitted to discuss with one another what it is, and decide on a single message that will be relayed down the line.[12]

Supervision: Like *Telephone*, but now there is someone overseeing the entire process, ensuring that the message has been relayed accurately and correcting participants if it hasn't been.

Each of *Double-Checking*, *Conferral*, and *Supervision* adds another element that is missing from *Telephone* and the earlier variations on it, in that they are so structured that errors are corrected not only by chance, but deliberately and in a way that is built into the transmissional practice. And once again, many religious traditions have elements that mirror these: similar to *Double-Checking*, those who transmit religious doctrine are usually around to witness how their teachings are conveyed by others, and intervene if things go wrong; similar to *Conferral*, it is possible for participants in a tradition to check with one another to ensure that the message is being received and transmitted accurately; and – more controversially, certainly, but centrally for someone like

[11] On the epistemic value of non-corroboration, see Coady, *Testimony*, pp. 213-214.

[12] Thanks to Angela Schwenkler for suggesting this variation.

Aquinas – similar to *Supervision*, the entire process is often thought to be governed by some kind of divine oversight that helps to eliminate errors. (Of course that may be mythical, but insisting on that point would be question-begging in this context.)

Finally, note that all of these modified *Telephone* scenarios can be combined, I think indefinitely: thus we could imagine a situation in which there are multiple 'first witnesses' who then create chains that diverge, backtrack, and converge; always double-checking, collaborating, and under the watch of a careful overseer; with participants who are highly reliable and motivated to get things right. The result of this will be a transmissional practice involving nothing more than chains of testimony and the oversight thereof, but whose outcome seems to be, as Aquinas suggests, at least as certain as most of the products of fallible human reason. If this isn't 'complete and accurate' knowledge, then very little of what we humans attain ever is.

Perhaps Zagzebski will reply that even if these cases show that reliance on the testimonial transmission of doctrine needn't render a religious tradition too epistemically fragile for its later members to have any religious knowledge at all, still there is *some* degradation that necessarily occurs when information is passed from one person to another. (Even if we imagine a *Supervision*-style case where the overseer is omnipotent, omniscient, and dead-set on ensuring the fidelity of the message, there may still be questions about whom we are to trust, and to what degree.) I am not sure about this, though: for one thing, in *Double-Checking* and *Conferral* the ability to confer with one's peers, plus the knowledge that one's predecessors have done the same, might justify those of us later on in the chain in being every bit as certain than the original witness or witnesses, if not more so. But here is one more variation that still fits the model of Tradition as Transmission, yet where those later on in the chain seem to be in an epistemically *superior* position to those who begin it:

> **Summation:** Like *Convergence*, but each original witness is given only a proper part of the message, which is put back together when the chains of testimony converge.[13]

The inspiration for *Summation* is the well-known parable of the blind men and the elephant: each one is touching some part of the beast

[13] For a similar case, see Coady's discussion of the 1983 Australian bushfires, in *Testimony*, p. 214.

and can report only what he feels; but we, who hear those reports, can combine them into an account of what they perceive that goes beyond each of their contributions. (I don't mean to say that the world's religions are like this.) Even if there has been some opportunity for each of their messages to be distorted before reaching us, still it could be that the sum of those messages tells us more about what they witnessed than any one of them heard in isolation.

I conclude that the model of Tradition as Transmission can account for the cross-generational stability of religious knowledge, contrary to Zagzebski's argument. But still it strikes me as insufficient, for reasons I expect Zagzebski will agree with. I turn to these in the final section.

III. INSULARITY

By my lights, the real problem with a pure model of Tradition as Transmission is not that testimonial transmission is too epistemically unreliable to extend religious knowledge to subsequent generations, but that the model is too *insular* in the way it envisions the development of a religious tradition and the role of faith in the life of religious believers.

The first way in which the pure model of Tradition as Transmission embodies an overly insular conception of faith is that it fails to recognize how knowledge from 'outside' a tradition can interact with the knowledge that is transmitted by it. To see this, consider one more imaginary case, now different from *Telephone* in a more fundamental way than the earlier variations on it:

> **Background:** A person is in receipt of knowledgeable testimony that *p*. She also knows other things relevant to the subject, which helps to situate *p* in a broader context. Thanks to this background knowledge and her willingness to situate what she learns within it, she becomes even more knowledgeable about this matter than others before her who have testified to it.[14]

As with *Summation*, the case of *Background* is one where the epistemic situation of a person later on in a chain of testimony is epistemically superior to the positions of those nearer the source. In this case,

[14] Similarly, Coady (*Testimony*, p. 216) suggests the possibility of using archaeological evidence to demonstrate the reliability of an oral tradition. His point there is somewhat different than mine, however. He comes closer to describing a *Background*-style case with his discussion of Kit Carson and the Indians on p. 219.

however, this is not because she is in receipt of *more* testimony than her predecessors, but because she does not rest content with what she is told, instead combining it with other things she knows to yield a level of understanding superior to that conveyed by the tradition alone. All this is possible because in contrast to *Telephone* and its variants (including *High-Stakes*, where transmissional accuracy is the only goal), the person described in *Background* is concerned not just with discerning and conveying a given message, but also with getting at the *truth* of what the message is about. (As Coady puts it, in taking the word of a witness and passing it along to another person it is essential that you treat the message 'as a worthwhile contribution to settling some issue.')[15] And as Aquinas writes in discussing the value of philosophical argumentation in theological matters, if truth is our concern then it should not matter where it comes from:

> ... the gifts of grace are so added to nature that they do not destroy it, but rather perfect it; so too the light of faith, which is infused in us by grace, does not destroy the light of natural reason divinely placed within us. And although the natural light of the human mind is insufficient to manifest the things made manifest by faith, still it is impossible that those things which have been divinely taught us through faith should be contrary to what has been placed in us by nature. For one of them would have to be false, but since both come to us from God, God would have to be the author of falsity, which is impossible. (*On Boethius' De Trinitate*, q. 2, a. 3, c.)[16]

Aquinas' immediate concern in this passage is to defend the use of philosophical texts and arguments as means to defend the faith and correct those who deviate from it, but his own theological writings are clearly in the spirit of *Background* as well: he treats the 'light of natural reason' as a source of knowledge of the material world and its creator, and draws constantly on philosophical and scientific concepts to extend his knowledge beyond what is simply conveyed to him by his tradition. In this way he accepts what has been transmitted to him by his forebears but also improves on that tradition from within, just as Augustine and other Christian thinkers had done before him.[17]

---

[15] *Testimony*, p. 220.

[16] In Aquinas, *Selected Writings*, ed. Ralph McInerny (New York: Penguin, 1998).

[17] Compare Alasdair MacIntyre's description of Aquinas' project in relation to his Aristotelian and Augustinian roots, in *Three Rival Versions of Moral Enquiry* (Notre Dame: University of Notre Dame Press, 1990), chs. 5-6.

As I suggested above, part of what makes it possible for a healthy religious tradition to relate in this way to sources of knowledge that lie outside the tradition itself is the fact that its participants conceive of it not merely as a vehicle for transmitting a message, but also as a means of getting at some important *truths*. That these truths are seen as *important* truths helps such traditions to evade what I see as Zagzebski's more incisive criticism of the simple model of Tradition as Transmission:

> This model cannot explain how a religious tradition is transmitted without some additional elements. Why would it matter to *us* what a man called Abraham did, or that Moses had a religious experience in front of a burning bush if we are only the distant recipients of testimony about their contact with God? If what tradition passes on is a reconstruction of someone's experience a long time ago, it is hard to see it as anything more than a historical curiosity, and their written texts as anything more than artefacts of an ancient culture. Chains of testimony do not add up to a tradition in a sense that pertains to religious belief unless the content of the testimony bears on the future recipients of the testimony. (p. 193)

Clearly, the point Zagzebski is making here applies even if a chain of testimony is so structured as to convey its message with perfect accuracy from one end to the other: in order to see this chain as part of a tradition in any meaningful sense, we need to know why the message is important to its members, and what real-life questions they take it to help them settle. This is, once again, something that is obviously missing from the set-up in *Telephone*, where the content of the message has no bearing on what people do outside the context of the game. And as I have emphasized, it is precisely because of the vital importance of what a religious tradition conveys that such traditions are able to embody transmissional practices that can convey information reliably in a way that the arrangement in *Telephone* does not.

The remarks I have just quoted also identify a further way that the model of Tradition as Transmission is overly insular, namely that religious faith – at least of the 'living' variety, if indeed there is any other – is not *just* a matter of accepting the truth of certain doctrines, but also requires a more-or-less thorough integration of that attitude with the other aspects of one's life.[18] Because of this, religious traditions don't convey bodies of

[18] Thanks to Jon Buttaci for encouraging me to develop this point. My talk of 'acceptance' is deliberate: I have in mind a 'purely intellectual' attitude taken toward a proposition, without any affective component or immediate dispositions to act on

doctrine alone, but also numerous things that are supposed to aid in the practice of faith, such as liturgy and other communal rituals, traditional practices of prayer and private meditation, various kinds of music and visual art, stories of individuals whose lives were somehow exemplary, ethical codes and catalogues of important virtues, and so on. Separated from these elements, religious faith runs the risk of degenerating into mere attachment to vague ideals or to the past for its own sake, with no sense of its living relevance. And as Zagzebski notes, all this means that the recognition of a religious tradition as authoritative is not just a matter of evaluating the truth of its doctrines, but also of seeing how engagement with the community in its practical directives will inform one's life as a whole[19]. At the same time, anyone who *does* take the teachings of their tradition as more than a 'historical curiosity' is bound to regard them in the way Aquinas suggests, as truths that can stand in a mutually informing relationship with things that are known in other ways. And this would be unreasonable if what the tradition transmitted could not be counted as knowledge.[20]

---

this basis. On the distinction between acceptance and belief, see L. J. Cohen, 'Belief and Acceptance', *Mind*, 98 (1989), 367-389. On the centrality of affect to religious faith, see J. L. Kvanvig, 'Affective Theism and People of Faith', *Midwest Studies in Philosophy*, 37 (2013), 109-128.

[19] E.g. she writes: '... trust that a particular religious tradition puts one in the best position to get at the truth depends in part on trust that it contains the highest attainment of the human spirit in relation to God. But to think that, one must have nonepistemic trust in the tradition and would need to determine that the tradition has that quality by the fact that its teachings satisfy conscientious reflection upon one's total set of psychic states, not just one's set of beliefs' (Zagzebski, *Epistemic Authority*, p. 200). And again: 'The Church is more than a body with the authority to reveal truths of faith and morals. There are other natural desires ... which can be better satisfied by participation in a wisdom community than on one's own. These desires include the desire to know and do the good, to acquire not just knowledge, but understanding, to learn patterns of living and principles of action that result in a more integrated self, to be surrounded by grace and beauty, and to experience the delights of living among persons whose own pursuit of those ends enhances one's own. The authority of a community can be justified by a conscientious judgment that these desires will be more satisfied by participation in the community.' (ibid., p. 201)

[20] Thanks to Rich Cordero, Matthew Miller, and George Stamets for taking part in a reading group on EPISTEMIC AUTHORITY, and to Jon Butacci, John O'Callaghan, and Angela Schwenkler for helpful conversations concerning these matters.

BELIEVING ON AUTHORITY

MATTHEW A. BENTON

University of Oxford

Linda Trinkaus Zagzebski's *Epistemic Authority* (2012) is a welcome exploration of the relationships between, on the one hand, the so-called 'cognitive' notions associated with epistemology (particularly knowledge, belief, justification, reflection, and rationality), and on the other, the commonly classified 'affective' notions of trust, desire, emotion, and reliance. In particular, she argues that the connection between them arises from our recognized dependence, both practically and rationally, upon epistemic authority, whether that authority resides in our own cognitive faculties or emotions, or in others' faculties and expertise. Such epistemic authority applies not only to mundane empirical matters such as our immediate natural environment, but also to the domains of morality and religion.

There is much to commend in this book. Zagzebski's treatment of these issues is thorough, and admirable for its broad vision of uniting social epistemology with topics in moral and political philosophy as well as philosophy of religion. Here I will concentrate on three main topics. In §1 I present some challenges for her view of rationality as it relates to self-trust; in §2 I consider how her view of authority relates to some issues of epistemic authority in testimony; and in §3 I raise some difficulties for her treatment of epistemic authority as it relates to religious epistemology.

I. SELF-TRUST AND RATIONALITY

Zagzebski begins by considering the ways in which we often encounter cognitive 'dissonance', wherein we experience conflict amongst our

mental states, including beliefs, emotions, desires, and decisions (2012: 29). She treats this notion of dissonance as basic, and notes that 'Many times when there is dissonance, the self automatically adjusts by giving up one of the states that conflict' (2012: 30). Given this starting point, she gives us a working definition of 'rationality' thus:

> I think that the awareness of dissonance resolved without effort gives us our initial model of what rationality is. I say that because I think that rationality is a property we have when we do what we do naturally, only we do a better job of it. To be rational is to do a better job of what we do in any case – what our faculties do naturally. (2012: 30)

From here Zagzebski proceeds, following Foley (2001) and Alston (2005), to argue that realization of the fact that there is no epistemically non-circular argument for the reliability of one's faculties leads us, upon reflection, to put our trust in our cognitive faculties as reliable means of getting the truth. But for Zagzebski, such self-trust is not the result of realizing that we lack 'full reflective justification' (in Alston's phrase), only after which we *then* resort to trusting the cognitive faculties we could not non-circularly prove to be reliable; rather, self-trust in our cognitive faculties is pre-reflective, operative even before we assess the matter of whether we have any epistemic reason or justification for trusting them.[1]

Discovery that epistemic circularity must be involved in any attempt to justify our reliance on our cognitive faculties for getting the truth leads, Zagzebski thinks, to a feeling of dissonance, because we naturally desire to seek the truth, and as self-conscious and reflective beings we examine whether our faculties can be (non-circularly) shown to get us the truth. Though Zagzebski does not spell this out explicitly, the dissonance presumably comes from wanting something (full reflective justification) which, upon scrupulous reflection, we discover we cannot have. And to the extent that we believed, or assumed, that we could not acceptably trust our faculties without the wanted full reflective justification, we

[1] A reason for Zagzebski's view here is that she understands trust as a three-place relation – 'One trusts something for some purpose or in some respect' – where the state of trust combines epistemic, affective, and behavioural components: 'when I trust *x* for purpose *y*, (1) I *believe x* will get me *y*, (2) I *feel* trusting towards *x* for that purpose, and (3) I *treat x* as if it will get me *y*.' (2012: 36–37) (One obvious difficulty is that clause (2) contains 'trusting', even though clauses (1)–(3) appear to offer at least a 'first approximation' of an analysis of what it is for one to trust something for some purpose. Perhaps this is easily remedied by instead having clause (2) read thus: I feel *hopeful* that *x* will get me *y*.)

either must give up the desire for that strong a justification, or the belief that we must have it acceptably to continue trusting our faculties.

Self-trust is supposed to be 'rational' because it helps us resolve this dissonance:

> Is it rational to have self-trust after reflection [on the circularity worry]? That depends, of course, on what we mean by rationality, and whether it applies to all three components of trust ... I said above that I think of rationality in the broad sense of doing a better job of what we do naturally in the use of any of our faculties ... Reflective self-trust resolves the dissonance we have when we discover epistemic circularity, and that seems to me to be rational. It is rational to believe that my faculties are trustworthy for the purpose of getting the truth; it is rational to treat my faculties as if they will get me to the truth, and it is rational to feel trusting of them in that respect. (2012: 43)

It is unclear to me how we ought to take Zagzebski's application of 'rational', given her meaning for that term, to this particular instance of dissonance. If being rational is just doing a better job of what we naturally do anyway, then if we did (prior to reflection on the matter) trust our faculties as reliable at getting us the truth,[2] then continuing to trust them for this purpose after encountering the circularity worry would quite clearly be continuing to do what we do anyway. But is continuing in such trust doing a 'better' job at it? This is hard to say; for on the one hand, maybe perseverance in trusting our faculties to get us the truth upon discovering that we cannot have the epistemic justification we wanted for it is doing it 'better', for one trusts even without the rationale for doing so that one had hoped to find. But on the other hand, continuing to do something for which one discovers one lacks an epistemic justification is often taken to be problematic. Zagzebski writes that 'It is rational to believe that my faculties are trustworthy for the purpose of getting the truth' even though I cannot find the non-circular justification for it that I had been seeking; and in general, if I find myself believing that *p* while having no evidence or epistemic grounds supporting *p*, continuing to believe *p* is normally thought to be less than epistemically rational (if not outright irrational), as that term is normally used. In short, stipulating that 'rational' means *doing better at what we do anyway* doesn't help us gain purchase on why continuing to trust our faculties for delivering the

[2] That is, trust them *selectively*, in the environments in which we recognise them to be most truth-conducive.

truth, though pragmatically inescapable,[3] is something worth doing or something we (epistemically) *ought* to do.

A related worry is that the dissonance which might be felt by a reflective person upon encountering the circularity problem is one that in fact needs no resolution, and if this is the case, there is no work for reflective self-trust to do. Zagzebski concedes that 'Some forms of dissonance do not need to be resolved; we can get along well enough with the dissonance. This often happens with conflicting desires, or with a desire that conflicts with a belief' (2012: 31), and I've suggested above that it is the latter type of conflict that self-trust is supposed to resolve. But what if instead what we actually do quite naturally is simply accept that we must live with the dissonance, and ignore it? For one thing that we also do naturally is distract ourselves from the stressful facts of our existence: perhaps, with Hume, we resort to socializing and backgammon to take our minds off the dissonance that serious reflection can bring.[4] On Zagzebski's preferred idiom, this Humean method is 'rational', because it would be doing better what we do naturally; but this coping strategy has little to do with resolving, as opposed to *avoiding*, the dissonance.

Notice the difficulty which Zagzebski's understanding of 'rational' raises for her arguments against the person who wants to trust his own faculties more than those of others (2012: 53). If someone thought he had no obligation to treat everyone as trustworthy whom he believes to be trustworthy, simply on the grounds that he prefers to trust himself and not others (or perhaps: trust himself more than he trusts others), Zagzebski thinks this would be 'unreasonable' for the person who 'cares about truth': for he would be more trusting of himself and his own faculties simply because such faculties are his own. But crucially, Zagzebski cannot say that doing this would be 'irrational', for on her view of what makes something rational, doing so might well be rational.

II. AUTHORITY, BELIEF, AND TESTIMONY

How should we approach the connections between belief, authority, and believing another's testimony, that is, believing what someone tells us on their authority? A natural place to start notes that typically, we

[3] Fricker (2014: 179) uses this phrase.

[4] Hume, *Treatise* (1978), Bk. I, pt. 4, sect. 7, para. 9.

regard another as authoritative when we believe she has strong epistemic grounds for what she tells us, and in particular, when she *knows* the thing she tells us. Supposing she does know what she asserts to us, we arguably have all the epistemic reason we need to believe what she's told us; indeed, we value another's say-so in large part because that is a primary way by which we can gain knowledge, and as such, we tend to feel cheated when someone testifies in the absence of knowledge. Appropriately asserting or testifying that *p* may be understood thus as requiring *knowledge* that *p*, or at least some kind of epistemic condition, for that is the condition on which a speaker has the *authority* to assert that *p*.[5]

Zagzebski's approach to these matters differs greatly. She distinguishes first-personal deliberative reasons from third-personal theoretical reasons, and defends an account of epistemic authority entirely from the first-person perspective:[6] she says that 'What is essential to authority is that it is a normative power that generates reasons for others to do or to believe something preemptively', where 'preemption' is 'a distinguishing feature of authority from the subject's perspective ... A preemptive reason is a reason that replaces other reasons the subject has' (2012: 102).[7] Her 'Preemption Thesis for epistemic authority' is this:

> The fact that the authority has a belief *p* is a reason for me to believe *p* that replaces my other reasons relevant to believing *p* and is not simply added to them. (2012: 107)

Thus Zagzebski is primarily interested in what it is for a person to be, or to be treated as, epistemically authoritative *for me*. On Zagzebski's view, someone's epistemic authority for me is intimately related to 'my conscientious judgment[s]' that, if I believe what the authority believes rather than trying to figure out what to believe myself, I will be (i) more

[5] See Williamson (2000: chap. 11, esp. 257): 'One can think of the knowledge rule as giving the condition on which a speaker has the *authority* to make an assertion. Thus asserting *p* without knowing *p* is doing something without having the authority to do it, like giving someone a command without having the authority to do so.' For advances of this view, see Turri (2011), Benton (2011 and forthcoming), Buckwalter & Turri (2014), and Fricker (2014), among others.

[6] Following Joseph Raz's (1988, 2009) account of political authority.

[7] 'Believing what another person believes or tells me preemptively is parallel to doing what he tells me to do preemptively. In both cases what the authority does gives me a reason to believe or do something that replaces my other reasons relevant to the belief or act. The kind of reason authority gives me is what is essential to it' (Zagzebski 2012: 102).

likely to form a true belief (**JAB 1**), and (ii) more likely to form a belief that survives my conscientious self-reflection (**JAB 2**).[8]

Applying this view to testimony, Zagzebski endorses a 'trust model' of testimony and ties it to being justified in relying on another's authority:

> the trust model of testimony is one in which telling gives the recipient a deliberative reason to believe what the speaker tells her. Trust is irreducibly first personal because it is a reason only for the person who has it. ... When you tell me that *p*, you ask me to trust you, and if I accept your invitation to trust, *I* trust *you*. (2012: 130–131)

Trusting your testimony to me gives me a reason for believing 'that preempts my other reasons for and against believing' what you tell me (2012: 132). Justification Theses for the Authority of Testimony, similar to (**JAB 1**) and (**JAB 2**), are endorsed:

> (**JAT 1**) The authority of a person's testimony for me is justified by my conscientious judgment that I am more likely to satisfy my desire to get true beliefs and avoid false beliefs if I believe what the authority tells me than if I try to figure out what to believe myself.
>
> (**JAT 2**) The authority of a person's testimony for me is justified by my conscientious judgment that, if I believe what the authority tells me, the result will survive my conscientious self-reflection better than if I try to figure out what to believe myself. (2012: 133)

Zagzebski later shifts to plausible 'Third-Person' versions of **JAT 1** and **JAT 2** as a way of handling the fact that authority seems less subjective than her first-personal principles make it out to be (cf. my worry in fn. 8): 'my conscientious judgment' of **JAT 1** and **JAT 2** is replaced by 'the fact that' in the Third-Person principles (2012: 137–138). Yet she contends that the Third-Person versions are 'a natural consequence' of their first-person counterparts:

---

[8] Zagzebski (2012: 110). Her *Justification Thesis 1 for the Authority of Belief* (**JAB 1**) reads: 'The authority of another person's belief for me is justified by my conscientious judgment that I am more likely to form a true belief and avoid a false belief if I believe what the authority believes than if I try to figure out what to believe myself.' I have difficulty understanding the language of another's authority being 'justified by my conscientious judgment ...', for it makes it sound like whether they are authoritative depends on a normative condition made possible by my own judgment. It would also seem that the added 'and avoid a false belief' is redundant, since if the target belief is true, it will avoid being false (assuming the law of non-contradiction).

> The point here is that if I can justify to others my taking a belief on authority under certain conditions, they can justify to me that I should take a belief on authority under the same conditions ... It follows that the third-person justification of epistemic authority is a natural consequence of the first-person justification. (2012: 138)

But on the one hand, Third-Person **JAT 1** does not follow from (First-Person) **JAT 1**: the latter can be fulfilled, or used to justify one's own belief, even though the former is not fulfilled. This is because on (First-Person) **JAT 1**, one's conscientious judgment is what justifies one, *whether or not* the teller is in fact more reliable than oneself at delivering the truth; whereas on Third-Person **JAT 1**, what justifies one is the fact that the teller is more reliable at delivering the truth. And on the other hand, Zagzebski's direction of argument here seems to me to get things exactly backwards: I will conscientiously judge that someone can serve as authoritative for me (or someone else) precisely in the situation where I judge her to be epistemically authoritative period. The latter condition is fulfilled when I judge her to know the proposition she is telling me or someone else, and in ideal cases I judge that because she does know it. In less than ideal cases where I've conscientiously judged someone to be authoritative when in fact she was not (e.g., when she doesn't know what she tells me), we'll be inclined to say that it was reasonable for me to believe on her authority even though she *lacked* epistemic authority on that occasion. Zagzebski's account does not deliver these results, and is to that extent counterintuitive.[9]

Furthermore, it would seem that Zagzebski's account of epistemic authority cannot explain our ability to identify who is epistemically authoritative in some domain even when they do not serve as authoritative *for us*. If I know that *p* and I can discern that you also know that *p*, then in seeing you tell Jane that *p*, I can judge that you are epistemically authoritative with respect to *p*, and worthy of Jane's trusting your testimony (on the matter of *p*, at least). Your epistemic authority concerning *p* does not, it seems, depend then on whether I or anyone else trusts you pre-emptively; what matters is whether you know, or are positioned to know, or are in some other strong epistemic position with respect to *p*.

[9] Notice also that a deceived deceiver (someone who intends to provide me with false testimony, but mistakenly provides true testimony) lacks epistemic authority in my favoured sense, but plausibly fulfils **JAT 1**. Thanks to Dani Rabinowitz for this example.

Finally, it does not seem essential to your possessing that epistemic authority that, if I believe *p* on your authority, my doing so replaces my own evidential reasons for *p*. Suppose I begin with some evidence *E* for *p*, which I regard as not very strong. When you tell me that you know that *p*, my decision to believe it on the basis of your authority need not replace my reasons generated by *E*. Indeed, my having *E* in the first place enables me to view your testimony as confirming what *E* weakly supported, namely that *p*; and in the right kind of case, part of my reason for trusting you as authoritative on this matter might be precisely that your testimony accords well, and perhaps explains, the evidence *E* I already have. Indeed, if *E* is in fact decisive evidence for *p* but I do not appreciate this, then if your testimony that *p* includes information that helps me see how *E* confirms *p*, I may rely on your authority without it replacing *E*. (Another case: suppose I already know that Jack went up the hill. You then testify that Jack and Jill went up the hill. I can accept the conjunction that: [Jack went up the hill and Jill went up the hill] on your authority, even if doing so does not replace my earlier reason for believing that Jack went up the hill.) For all these reasons, it seems that the Preemption Thesis, at least for testimony, is implausible.

III. AUTHORITY AND RELIGIOUS EPISTEMOLOGY

In her chapter on religious authority, Zagzebski discusses how one can justify religious belief on the basis of others' religious beliefs, particularly given the *prima facie* reason available through a *consensus gentium* argument for theism from self-trust (2012: 185–188). More significant is her argument for believing on trust within communities (cf. also Chap. 7), especially communities of specific religious traditions; particularly incisive is her discussion of how such communities function as part of a religious tradition, and how the maintaining of, and the participation in, such a tradition over time depends on both the conception of divine revelation at work in such a tradition, which itself contributes to the structure of the tradition. The tradition's beliefs, its motivational, moral, and spiritual values, and its learned patterns of living are organized around the tradition's view of divine revelation: these components of the tradition reflect the tradition's view of how its participants' may learn from, or come into contact with, what God has revealed of God's self and God's purposes for them.

A common way to understand divine revelation (at least within the monotheistic traditions) is to think of God's revelation on the model of a kind of divine testimony to us.[10] One model emphasizes a chain of unbroken transmission from original historical sources to whom God gave the revelation, and the work of the tradition is to maintain and hand on that testimony to later generations who are the ongoing recipients of the testimony. Another prominent model emphasizes instead the current recipient's experience of God rather than solely the original revelatory experience (though typically the recipient's experience is in some way mediated by interaction with the preserved historical account of earlier divine revelations). On this model, if Scripture preserves some of the original revelation by way of testimony, it nevertheless 'speaks directly to the reader or hearer without any need for a tradition of interpretation of authority in its exposition and preservation' (2012: 194). On the Christian version of this model, the Holy Spirit enables this kind of first-hand contact with God: a person can, by the grace of the Holy Spirit's work, come to (or deepen one's) faith through receiving the Gospel proclaimed. (These are not the only two models, but such models are characteristic of many traditions, even within strands of a particular religion.) On either view, the tradition serves to preserve and interpret the divine testimony over time, and to shape its participants given the model of divine revelation with which it operates.

Zagzebski argues that her approach to authority can serve as an important justifier for the religious believer insofar as she has argued that such a believer can justifiably trust the authority of the tradition. She expresses dissatisfaction with recent religious epistemology which focuses too much on either first-hand experience or on the chain model of testimony: for

> I can trust my tradition more than my own experience in many cases, and of course my experience is limited to the experience of one person. Given that we reasonably take beliefs from others or based on the experience of others, the structure of the process by which those beliefs are dispersed within a community and continued through the future

[10] Aquinas's view of revelation (in *Summa Theologiae* IIaIIe, Qu. 6. Art. 1), whereby one's will with a divinely inspired inclination moves the intellect to accept primary truths of faith, is arguably non-testimonial; cf. Hawthorne (2013, esp. §3). Another noteworthy exception is Maimonides' non-testimonial account; see Rabinowitz (2013: Ch. 2, esp. 82ff.), as well as Stern (1998); for similar accounts in the Islamic tradition, see Davidson (1992).

> life of the community needs epistemological models. ... I am suggesting a rule of justification that bypasses the chain model ... (2012: 202)

Furthermore, she is dissatisfied with the evidence view of testimony (2012: 128–131) particularly as applied to divine revelation: 'Religious faith is impossible to explain, much less justify, on the evidence view of testimony. That view forces us to either redefine faith as belief on a certain kind of evidence, as Locke did, or we must say that faith is non-rational, based on emotions that have nothing to do with epistemic justification' (2012: 202). Having dispensed with first-person experience, with chain models of divine revelation, and with the evidence view of testimony, Zagzebski clears the way for her trust model of divine testimony operative within a communal tradition.

While I don't disagree with some of her reasons for dissatisfaction here, I do not think that we are forced to choose between a trust model and an evidence model when it comes to divine testimony; nor does it seem right to say that religious faith is impossible to explain, or justify, on the evidence view. Religious faith may be understood in terms of evidence; Lara Buchak (2012 and 2014) has offered an account of having faith (expressed through action) where such faith can be rational given its relation to evidence. Moreover, insofar as a tradition decrees as sacred texts which are thought to document some original divine revelation, those texts form a portion of the divine testimony that may be evaluated by historical standards of evidence (to say nothing of evaluating such documents for authenticity). Finally, endorsement of a trust model of divine testimony may bring with it a concern for evidence because one will regard oneself as trusting the whole of a tradition for aid, at times, in determining what exactly the *content* of divine testimony is: for example, I must evaluate my own tradition's claims about who God is, what demands God may make of me, or what God may be trying to teach me, when attempting the (communal, not merely individual) process of discerning what God is revealing (or has revealed) to us or to me. This requires weighing evidence about my tradition's trustworthiness on such matters, including the evidence that the tradition's resources may underdetermine exactly what, and how, God is communicating to us presently.

There is a more fundamental worry, however. On Zagzebski's trust model of testimony, S's telling you that p invites you to trust S regarding p, and when you accept that invitation, you believe p on S's authority.

But this model seems to assume that one knows who – namely, S – is telling you that *p*. A major difficulty with applying this model to divine testimony is that the believer must believe, or take on faith, that what has been testified to her really is from God (what if it is in fact generated subliminally by her own self-interests?). The first-personal deliberative nature of the reason for acceptance is lost if one is in serious doubt about the source of the testimony. Even if one is confident of God's existence, the process of discerning whether some seemingly divine testimony – be it a recent insight, spiritual directive, theological interpretation of Scripture, etc. – is really from God can be a difficult epistemic task.[11] And it seems to me that this epistemic task cannot be separated from the relevance of evidence, including how the testimony of Scripture, its interpretation in one's tradition, and the testimony of spiritual exemplars provides a kind of evidence for how one ought to evaluate (purported) divine testimony. But even once one satisfies oneself that some revelation is from God, Zagzebski is right that the invitation to trust remains; and in the divine case, one's ability to trust God concerning such testimony, and the outcomes of acting upon it, is part and parcel of what it is to have faith in God.[12]

BIBLIOGRAPHY

Alston, William P. 2005. *Beyond 'Justification': Dimensions of Epistemic Evaluation* (Ithaca: Cornell University Press)

Benton, Matthew A. 2011. 'Two More for the Knowledge Account of Assertion', *Analysis*, 71: 684–687

Benton, Matthew A. Forthcoming. 'Gricean Quality', *Noûs*

Buchak, Lara. 2012. 'Can it Be Rational to Have Faith?', in Jake Chandler and Victoria S. Harrison (eds), *Probability in the Philosophy of Religion* (Oxford: Oxford University Press), pp. 225–247

Buchak, Lara. 2014. 'Rational Faith and Justified Belief', in Laura Frances Callahan and Timothy O'Connor (eds), *Religious Faith and Intellectual Virtue* (Oxford: Oxford University Press)

[11] Though sometimes it may be immediately clear: see Wolterstorff (2010: 317-322) for discussion of such a case.

[12] Thanks to John Hawthorne, Dani Rabinowitz, Wes Skolits, and an anonymous referee for helpful feedback. This publication was made possible through the support of a grant from the John Templeton Foundation. The opinions expressed in this publication are those of the author and do not necessarily reflect the views of the John Templeton Foundation.

Buckwalter, Wesley and Turri, John. 2014. 'Telling, Showing, and Knowing: A Unified Theory of Pedagogical Norms', *Analysis*, 74: 16–20

Davidson, Herbert. 1992. *Alfarabi, Avicenna, and Averroes on Intellect* (Oxford: Oxford University Press)

Foley, Richard. 2001. *Intellectual Trust in Oneself and Others* (New York: Cambridge University Press)

Fricker, Elizabeth. 2014. 'Epistemic Trust in Oneself and Others – An Argument from Analogy?', in Laura Frances Callahan and Timothy O'Connor (eds), *Religious Faith and Intellectual Virtue* (Oxford: Oxford University Press), pp. 174–203

Hawthorne, John. 2013. 'Aquinas on Faith and Knowledge: Response to Robert Pasnau', in John Marenbon (ed.), *Continuity and Innovation in Medieval and Modern Philosophy: Knowledge, Mind, and Language* (Oxford: Oxford University Press)

Hume, David. 1978 [1739–40]. *A Treatise of Human Nature*, 2nd edition (Oxford: Clarendon Press)

Rabinowitz, Dani. 2013. *Knowledge by Way of Prophecy*, Ph.D. thesis, Philosophy Faculty, Oxford University

Raz, Joseph. 1988. *The Morality of Freedom* (Oxford: Clarendon Press)

Raz, Joseph. 2009. *Between Authority and Interpretation* (Oxford: Oxford University Press)

Stern, Josef. 1998. *Problems and Parables of Law* (New York: SUNY Press)

Turri, John. 2011. 'The Express Knowledge Account of Assertion', *Australasian Journal of Philosophy*, 89: 37–45

Williamson, Timothy. 2000. *Knowledge and its Limits* (Oxford: Oxford University Press)

Wolterstorff, Nicholas. 2010. 'On Being Entitled to Beliefs about God', in Terence Cuneo (ed.), *Practices of Belief: Selected Essays, volume* 2 (Cambridge: Cambridge University Press), pp. 313–333

Zagzebski Trinkaus, Linda. 2012. *Epistemic Authority: A Theory of Trust, Authority, and Autonomy in Belief* (Oxford: Oxford University Press)

INCONSISTENCY, UNCERTAINTY AND EPISTEMIC AUTHORITY

DAMIAN LESZCZYŃSKI

University of Wrocław

The book of Linda Zagzebski deals with the serious and complex problem of epistemic authority – a problem concerning not only epistemology, but also some topics in practical philosophy: ethics and political philosophy. It is hard to refer to all of the issues brought up in the book, so I will try so to concentrate on those which, in my opinion, are most essential and involve the problem of self-reliance. This issue is essential and its correct analysis is a condition of the possibility of conducting further argumentation. I do not intend to summarize all solutions presented in the book; I am only going to refer to those aspects of Zagzebski's argumentation which are, in my opinion, problematic and doubtful: the problem of dissonance and its elimination as rational purpose; the problem of 'naturalness', the reconstruction of epistemic authority and, finally, the interpretation of Descartes' philosophy.

I.

Zagzebski's statement that our thinking consists in, generally speaking, the elimination of dissonance (that is the incompatibility or conflict between our acts or mental states), is similar to the well-known conception of Charles S. Peirce which was introduced in the article *How to Make Our Ideas Clear*. However, it means that analyses have a rather psychological character and not epistemological (in Peirce's philosophy, similarly as in the philosophy of Dewey, epistemology was in fact replaced by cognitive psychology). But in this way we are going from the level of *questio iuris* to the level of *questio facti*, losing the normative dimension of our investigations which seem essential when we try to

solve problems concerning morality (discussed in chapter 8). It seems that we should first consider what in fact is a priority for us: harmonious views (psychological optimum) or whether they are true (epistemological optimum)? It is an important issue because these things are not necessarily connected with each other. It is not that when we achieve harmony and coherence (as a result of the elimination of dissonance) we automatically achieve the truth. It can simply be the other way round. It is often the case that we achieve harmony at the cost of the truth, for example in various forms of ideology or pseudoscience (e.g. David Icke's theory of reptilian humanoids is perfectly cohesive, but it seems at the same time totally false).[1] I would even say that the excellent cohesion of certain systems of statements or opinions should arouse our suspicion: e.g. some schizophrenics' systems of beliefs are truly harmonious and their verification from the point of truth must lead to the introduction of a deep dissonance. Also in the history of science we have numerous examples of the introduction within existing conceptual schemes of some hypotheses *ad hoc* in order to provide them with harmony (e.g. attempts to rescue Newton's mechanics in the late 19th century). However, it did not increase the cognitive power of these schemes and ultimately did not inspire greater confidence.

II.

I have great doubts concerning whether aiming to remove dissonance can be regarded *en bloc* as a rational action. Nowadays a similar position is taken by Jurgen Habermas, who proposed a theory of communicative actions as a general theory of the equalization of dissonances. His idea of a harmonized communication based on a global *consensus*, is presented as an *a priori* rational purpose. However, as presented by, among others, Nicholas Rescher in his book *Pluralism: Against the Demand for Consensus*, this assumption is highly controversial. Rescher's view is exactly contrary to Habermas's – both in ourselves, as well as within a community, a conflict, a dispute and dissonance are rational because they have a stimulating effect. Moreover, from the point of

[1] Coherence may be a necessary condition of truth in the case of sets of statements but not in the case of an individual statement. We can imagine a set of incoherent statements which are individually true, for example pragmatically. It is one of the consequences of Gödel's incompleteness theorems. (This, however, is not the place to discuss this problem in detail.)

view of Popperian falsificationism (which in some respects is a reliable description of how a scientific community works) a rational action is the pursuit of dissonance and the introduction of an inconsistency into the system of our beliefs by submitting our hypotheses to extremely restrictive tests. Let us also recall a classical example of Socrates. His method consisted exactly in the demolition of harmony, the introduction of dissonance, and finally providing his audience with no clear answers but with new questions and problems. I would say that this is the core of any philosophical and intellectual activity: posing questions, formulating doubts, introducing dissonance and formulating questions. And the fact that in solving fundamental philosophical problems, in my opinion, no significant progress has been made since the time of Plato, shows that a dissonance is included in the nature of philosophical enquiry and is something which supports it. So it seems that we do not need to eliminate dissonance but we have to maintain it.

III.

In this context it is possible to make some classifications of dissonances and their forms. Zagzebski makes some classification in her book (p. 50) where she mentions some dissonances between beliefs, emotions and actions as well as dissonances between beliefs and desires. I think that these divisions can be more sophisticated and systematic when we consider dissonance or incoherence on the logical or syntactic level (between statements), epistemic (between cognitive acts and their objects, i.e. between what I see and what is), as well as pragmatic (between statements and acts). It is evident that there is a difference between immanent dissonance, taking place inside the subject between his mental states (cognitive, volitional, emotional), and transcendent dissonance between the subject (and his mental states or acts) and the world (that means: objects, persons, acts, cultural values, etc.). Considering different types of dissonances, it is clear that even acknowledging that rational action consists in the elimination of dissonance, it appears there is no one way of this elimination, because a contradiction of my opinions is something different than a conflict between acts and values or acts and opinions. In my opinion different types of dissonance require us to take into account different types of the rationality that would enable us with regards to them. Subsuming everything under one general category of rationality is not a good idea because excessive generality lacks its

predicate value. One should also consider certain meta-rationality which is responsible for choosing some local rational actions in reference to specific types of dissonance.

IV.

I am doubtful about the idea of the 'normality' and the 'naturalness' used by Zagzebski (e.g. pp. 86, 201-202, 251-254, where 'normal' is tied together with 'harmony'). It is hard to state whether these terms are used normatively or descriptively, and whether they refer to a biological, cultural, or social norm, or are understood psychologically and associated with the internal integrity of the individual. And since Zagzebski closely connects the idea of rationality with ideas of the normal and the natural, we encounter a big problem here, especially that the concept of rationality – as I demonstrated above – is also problematic. In this context an attempt to solve some sceptical problem by referring sceptical doubts to 'what we naturally do' seems fallible (p. 45). It is not only due to the vagueness of this notion, but also because it is not known why the 'naturalness' would be actually regarded as a criterion of approval in case of philosophical scepticism (I will come back to this issue at the very end of this text).

V.

I have some objections to the very analysis of the phenomenon of epistemic authority. This analysis is based on some thoughts of Joseph Raz (p. 106), concerning political authority and is some kind of extrapolation of his political ideas in the field of cognition. In my opinion it cannot be done easily, because the nature of political authority is connected with action (practice) and so differs from the nature of epistemic authority, associated rather with speculation and theory. The first authority concerns *techne*, the second *episteme* (or *doxa*) (it is possible here to refer to Oakeshott's classical division of practical knowledge, that is ability, and theoretical knowledge – knowledge in the proper sense of the word). So I would disagree with the statement that epistemic authority 'has all of the essential features of practical authority' (p. 139), because these are two completely different kinds of the authorities. Therefore applying Raz's political analyses to the area of epistemology is unfortunate and cannot be beneficial but rather makes the core of the issue obscure.

I think that clear and formal characteristics of epistemic authority (and authority in general) proposed half a century ago by Joseph Bochenski, would be a good starting point here.[2] Bochenski wrote that authority is a relation where *p* is an authority for *q* in the domain of *D*, when *q* accepts everything what *p* is offering him, and what belongs to the domain of *D*. Bochenski precisely studied cases of justified and unjustified use of authority, trying to show that it is possible to describe them from an external point of view (third-person perspective), not only from the point of view of an inner conviction of an individual (because there is often a case when a deep confidence of the individual in the authority is unjustified). Bochenski analyses not only conditions in which somebody can be an epistemic authority for someone, but also general conditions for being an authority for someone in a certain domain. The central question for Bochenski is the distinction between epistemic and deontic authority corresponding to the above discussed distinction into *episteme* (theoretical knowledge) and *doxa* (practical knowledge). However, according to my interpretation, the most important is Bochenski's statement that in the domain of practice – that is in morality, politics, the art of war and the like – there is no (and cannot be) epistemic authority. It results from the fact that there is no scientific theoretical knowledge which would enable anyone to become an epistemic authority in moral, political or existential issues. It is possible to be an epistemic authority in a field of science of morality or science of politics – but not in morality itself or politics itself. In these domains there are of course some authorities, but they have a completely different, deontic character. It shows that applying a structure of political (practical) authority to epistemic problems cannot be made. Obviously, the thesis about the non-existence of epistemic authority in the domain of practice can be questioned, for example proponents of ethical intellectualism would surely disagree with it, but regardless of that I think that Bochenski's analysis deserves attention.

VI.

The last matter which I would like to discuss here is connected with a way in which Zagzebski analyses the philosophy of Descartes. Admittedly, in her book we find only brief references to this thinker in the context of the problem of self-reliance. However, I think that it is worthwhile to devote

[2] Joseph Bochenski, *Was ist Autorität?* (Freiburg: Herder, 1974).

more attention to him because the problem of the relation of self-reliance and epistemic authority is precisely analyzed in Descartes' writings.

Let us start with a simple issue. Zagzebski is right on one hand when claiming that 'it is a mistake to interpret either the Cartesian method of doubt or his foundationalism as a justification of self-reliance' (p. 17) (the problem is more complicated and I will refer to it later on). However, on the other hand she is wrong when stating that Descartes' object of trust is 'the use of human power purified by his method' (ibid.). Indeed, Descartes writes in several places in *Meditations* that everything that we clearly and distinctly understand is true, but this statement is conditional. The method itself is not trustworthy, but needs a completely external guarantee. And this guarantee is divine truthfulness. In *First Meditation* Descartes analyzes cases of clear and distinct cognition, starting from sensory perception, and finishing on *a priori* mathematical reasoning, showing that each of these types of cognition can be dubious. Not only senses can deceive me – I can be deranged, I can dream – but it is also possible that God still deceives me (the hypothesis of the *deus deceptor*) or there is an all powerful demon deceiver who provides me with an entire image of my world and influences my will in such a way that I accept falsehood for the truth, e.g. when I think that 2 + 3 = 5 (the hypothesis of *malin genius*). Certainly, even if the *malin genius* exists and deceives me constantly, at least I am sure about my own existence, but nothing more, much less the existences of any external world. Therefore Descartes must prove that God is not a deceiver (although he needs to show first that I am not in the power of the almighty demon). This proof, as is widely known, is made in *Third Meditation* through the so-called ontological argument (and we know that this argument encounters substantial criticism). Descartes presents additional arguments in *Fifth Meditation*, where he writes: 'But once I perceived that there is a God, and also understood at the same time that everything else depends on him and that he is not a deceiver, I then concluded that everything that I clearly and distinctly perceive is necessarily true.'[3] So one can see that for Descartes the authority of the method is absolutely relative towards the authority of God and the authority of God is justified when I know that God is not deceiver.

[3] René Descartes, *Meditations, Objections, and Replies*, Edited and Translated by R. Ariew and D. Cress (Indianapolis/Cambridge: Hackett Publishing Company, Inc., 2006), p. 39.

Therefore we clearly see that in the context of the philosophy of Descartes we cannot identify 'epistemic self-reliance with epistemic autonomy', as is suggested by Fricker and quoted by Zagzebski (p. 18). The situation is exactly the opposite: I can trust myself only because I am not autonomous – because I am God's creation and as such I am not misled by God.

But what happens in a situation if God is not my creator, if I was created in some other way? This eventuality is almost universally accepted by contemporary naturalized philosophy. Descartes also analyses this problem:

> Perhaps there are some who would rather deny so powerful a God, than believe that everything else is uncertain. Let us not oppose them; rather, let us grant that everything said here about God is fictitious. Now they suppose that I came to be what I am either by fate, or by chance, or by a connected chain of events, or by some other way. But because deceived and being mistaken appear to be a certain imperfection, the less powerful they take the author of my origin to be, the more probable it will be that I am so imperfect that I am always deceived. I have nothing to say in response to these arguments. But eventually I am forced to admit that there is nothing among the things I once believed to be true which it is not permissible to doubt – and not out of frivolity or lack of forethought, but for valid and considered arguments. Thus I must be no less careful to withhold assent henceforth even from these beliefs than I would from those that are patently false, if I wish to find anything certain.[4]

So here we have a clearly posed problem: if I am not created by a good God, if I am not cognitively heteronomous on the epistemic level (e.g. if the basis of my cognition is not dependent on God), I cannot have certain knowledge and I can never trust myself (I cannot believe even in the above statement). If I am a contingent creation of nature, it is possible that I am able to know something truly, but I am not able to verify it – I cannot know that I know something (in other words: even if I accidentally find the truth, I cannot know the criterion to distinguish truth from falsehood). But there is also a possibility that I was created by the evil demon deceiver, and it means that I will never find the truth. But of course this statement is self-referential and can also be false. In this case we finally reach a paradox.

[4] Ibid., pp. 11-12.

However, Descartes' solution to the problem of self-reliance is more complex. He writes that we cannot trust ourselves and our knowledge because our senses and our own minds can deceive us. But at the same time Descartes seems to trust himself when he states that senses can deceive him and the mind can be wrong. Therefore here we are faced with two levels on which the issue of the self-reliance is being considered: empirical and transcendental. The first deals with what is called objective knowledge, the act of cognition in which the mind is directed towards transcendent objects (or objects that are recognized as transcendent). Here I should not trust myself, because senses can deceive me, the mind can draw wrong conclusions, etc. But the second level deals with a type of meta-knowledge, meta-cognition where the mind is directed towards itself as the subject who gets to know an object. The discourse of *Meditations* is conducted at this very level: the subject (*ego*) has become an object of examination. And in this case Descartes seems to trust himself: when he is formulating sceptical arguments from the *First Meditation* he believes that he can draw a conclusion about the impossibility of all objective knowledge. But is he right to do so? Can that meta-knowledge, which may be recognized as a kind of transcendental knowledge, be also deceptive and wrong? Can it not fall under methodic scepticism and sceptical arguments? It is certainly possible. My impressions concerning external things can be wrong in the same way as my knowledge about myself and about my own cognition. And if on the level of the meta-knowledge I find the idea of God which would be a guarantee of the certainty of objective knowledge, this very act can also be doubtful. Perhaps the idea of God and his goodness, which implies the impossibility of deception, is only the result of the evil demon's operations who in this way is trying to deceive me about the value of the clear and distinct cognition.

The conclusion is that after all I could trust myself in the act of objective (empirical) knowledge only if I could trust myself in the act of meta-knowledge (transcendental knowledge) which determines the methodological acceptance criteria of statements which appear at the first level. In other words, I could have a real objective knowledge if I could correctly apply the method, but the method is reliable only when God is not a deceiver and when we were not created by accident. However, in order to know that I was created by a good God, I must already rely on the method applied at the meta-level. But this falls under

the hypothesis of the evil demon who can mislead me into thinking that I should not be certain about conclusions formulated purely apriorical and analytically. After all it seems that I cannot go beyond myself – I have no authority except me, but my authority is confined to the certainty of my own being – independently of its genesis, nature and purpose. And it is necessary here to add that it is not even self-knowledge, but rather a kind of elementary self-awareness, a primitive intuition accompanying all of my mental acts (and also cognitive mistakes). It is hard to go beyond this intuition and it is hard to get something out of it.

VII.

It may seem that this last issue goes beyond any problems discussed in Zagzebski's book. But it does not. In this example we can see that the problem of the possibility of self-reliance, and of basing it on a self-knowledge, is closely related with the question about my genesis. It is possible to say that the three fundamental questions of Kant – what can I know? what ought I to do? what may I hope? – depend on an answer to this question: where do I come from? And so it seems that the issue of self-reliance is a little bit more complicated than it was presented by Zagzebski and cannot be easily solved by appealing to some of our natural and normal abilities or the common-sense attitude. In my opinion the only way to correctly grasp and formulate this issue is to start from the position of Descartes, or, generally speaking, from the transcendental position, parenthesize our natural attitude, that is everything that common sense and the positive sciences proclaim. I think this is how real philosophy should work. Solving philosophical problems by appealing to common sense is, unfortunately, mixing two diametrically different types of the discourse – philosophical reflection, on the one hand, and our natural attitude on the other. From a philosophical point of view such answers (e.g. common sense solutions to philosophical questions) are trivial, whereas from a common sense perspective they are unnecessary and redundant, because common sense deals with them without any explanations. Kant wrote in *Prolegomena*: 'To appeal to ordinary common sense when insight and sciences run short, and not before, is one of the subtle discoveries of recent times, whereby the dullest windbag can confidently take on the most profound thinker and

hold his own with him. So long as a small residue of insight remains, however, one would do well to avoid resorting to this emergency help.'[5] I think that his opinion still prevails.

[5] Immanuel Kant, *Prolegomena to Any Future Metaphysics That Will Be Able to Come Forward as Science*, trans. G. Hatfield (Cambridge: Cambridge University Press, 2004), p. 259.

CONSCIENTIOUS SELF-REFLECTION TO THE RESCUE?

JOSHUE OROZCO & NATHAN L. KING

Whitworth University

Linda Zagzebski's *Epistemic Authority*[1] is an impressive, important, and wide-ranging book. We find much to admire within its pages. But in the spirit of philosophical interaction, our goal is to foster discussion of issues concerning which we find Zagzebski's treatment less than wholly compelling. We focus on (i) Zagzebski's assessment of the recent disagreement debates; (ii) the role of conscientious self-reflection in her solution to the epistemic problem of disagreement; and (iii) the broader role of conscientious self-reflection in her project. We argue that Zagzebski's notion of conscientious self-reflection is neither necessary nor sufficient for rational belief; nor does it provide the sort of cognitive guidance that is claimed for it. These considerations, we think, call for further clarification regarding the central role that Zagzebski gives to conscientious self-reflection. They thereby leave in doubt her specific solution to the problem of disagreement.

THE PROBLEM OF DISAGREEMENT

Suppose you believe that *p*. You then come to learn that someone as reasonable and conscientious as you are, and who has the same relevant evidence you have, disagrees with you about *p.* How should you

[1] Linda Trinkaus Zagzebski, *Epistemic Authority: A Theory of Trust, Authority, and Autonomy in Belief* (New York: Oxford University Press). References to this book will be bracketed within the text.

respond? Recent discussion of this question has given rise to a spectrum of views. At one end of the spectrum are views on which learning of such disagreement is epistemically irrelevant. At the other end are views on which learning of such disagreement carries great epistemic significance – under such conditions, one should give equal weight to one's dissenters and abandon the belief that *p*.[2]

Zagzebski's diagnosis is that views on both ends of the spectrum stem from two competing and distinctively Modern values: the ideal of *self-reliance* (*egoism*) and intellectual *egalitarianism*. Views that accord no epistemic weight to a dissenter's opinion are supported by the ideal of self-reliance. According to such views,

> It is *my* conscientiousness that counts for me, not hers. I trust *my* reasoning and other epistemic powers and not hers because my powers are mine and hers are hers. (p. 204)

In contrast, views that assign very significant weight to dissenting opinions are supported by egalitarianism:

> When my belief conflicts with the belief of another person whom I judge has epistemic powers and virtues equal to mine, there is no reason to think that I am the one who is right, and so I have no reason to keep my belief. (p. 205)

Consider, then, the following positions regarding disagreement, both of which Zagzebski regards as 'extreme':

> *Egoism:* In the face of peer disagreement, one should give one's dissenter's opinions no epistemic weight (and thus retain one's belief with unaltered confidence).
>
> *Egalitarianism:* In the face of disagreement, one should give one's dissenter's opinions weight equal to that of one's own (thus moving from, say, belief to suspending judgment).

---

[2] For a seminal defence of the view that awareness of peer disagreement carries no independent epistemic weight, see Thomas Kelly, 'The Epistemic Significance of Disagreement', *Oxford Studies in Epistemology*, 1 (2005), 167–96. For seminal defences of the view that awareness of such disagreement is epistemically weighty, see Richard Feldman, 'Epistemological Puzzles About Disagreement', in Stephen Hetherington (ed.), *Epistemology Futures* (New York: Oxford University Press, 2006), pp. 216–36; David Christensen, 'Epistemology of Disagreement: The Good News', *Philosophical Review*, 116 (2007), 187–217; and Adam Elga 'Reflection and Disagreement', *Noûs*, 41 (2007), 478–502.

Zagzebski finds both options unpalatable. On the one hand, egoism seems to implausibly ignore – and treat as irrelevant – the fact that others are our epistemic equals. On the other hand, egalitarianism seems to lead to scepticism for many of our beliefs, and inappropriately ignores the unique role of the *self* in the resolution of epistemological problems. So what should we do? 'The commonsense solution', Zagzebski says, 'is to compromise. Why not say that we should not place too much more trust in our own faculties than in those of others, but that we need not go so far as to think of others as our equals?' Unfortunately, she thinks, this initially attractive solution is 'actually hopeless' (p. 206). She claims that while there is a principle argument for both of the extreme positions, 'There is no argument for a compromise position on trust other than the desire to avoid the two extremes' (p. 207). Indeed, she thinks, a middle-ground between the two extreme options is likely to be theoretically *worse-off* than the extremes themselves, because it may not itself enjoy support from either extreme.

Zagzebski is of course aware that middle-ground views exist. One such view is the *total evidence view*. On this view, the rational attitude to take with respect to some proposition in the face of disagreement is fixed by one's total evidence, where this includes relevant *first-order evidence* (evidence directly relevant to the target proposition) and *higher-order evidence* (evidence about our evidence, or about our capacities for or performance in responding rationally to our evidence).[3] Thus, on total evidence views, the attitude that is rational to hold in the face of disagreement will depend both on the evidence that is directly relevant to the disputed proposition *and* evidence concerning the disagreement itself.

Crucially, advocates of the total evidence view attempt to give principled reasons for a compromise between egoistic and egalitarian values. Note that the character of the relevant first-order evidence makes no appearance in either of the extreme views. According to egoism, all that matters is my conscientious judgment; and this judgment matters because it's mine. On egalitarianism, all that matters are the opinions of epistemic agents – namely, myself and my dissenter. But according to the total evidence view, it is a mistake to neglect first-order evidence

[3] For a defence of the total evidence view, see Thomas Kelly, 'Disagreement and Higher-Order Evidence', in Richard Feldman and Ted A. Warfield (eds.), *Disagreement* (New York: OUP, 2010), pp. 111-74.

in these ways. Indeed, in the absence of disagreement, such evidence is *paradigmatically* what makes our beliefs rational. It would be surprising if its relevance simply vanished once disagreement reared its head. By focusing on such evidence, in addition to higher-order evidence, we can see how total evidence views offer a principled middle way between egoism and egalitarianism. For perhaps in all cases, the higher-order evidence that someone one trusts believes $\sim p$ is epistemically relevant to some extent. But perhaps in some cases, the relevant first-order evidence supporting p is so strong that it makes belief that p more rational than denying or withholding – even when the higher-order evidence is accounted for.

It is initially unclear why Zagzebski is dismissive of middle-ground views, including the total evidence view – especially in light of her own solution to the problem of disagreement. That solution features prominently the notion of conscientious self-reflection: 'Given the argument of this book, it is reasonable to resolve the conflict in favor of what I trust the most when I am thinking in the way I trust the most, that is, conscientiously.' (p. 214) When thinking conscientiously, Zagzebski says, an agent considers both *theoretical reasons* (e.g., arguments), and *deliberative reasons*. Deliberative reasons are irreducibly first-personal reasons that connect *me* to the truth of p (p. 64). Such reasons include items like intuitions and experiences. Crucially, they can also include instances of trust in others (p. 65). But notice: these latter items are *higher-order evidence*: they are evidence about whom I should trust to have good evidence, and to evaluate evidence rationally. Thus, Zagzebski's view entails that the conscientious thinker will consult both first and higher-order evidence in the face of disagreement – a position that seems quite close to the total evidence view.[4]

The above remarks can be seen to yield a trilemma for Zagzebski: (i) accept egoism and inherit whatever problems attend it; (ii) accept egalitarianism and inherit whatever problems attend it; or (iii) accept that there is a principled middle way between egoism and egalitarianism – contrary to her initial judgment. Inasmuch as Zagzebski herself seems

[4] In this way her view is in agreement with total-evidence views. However, given the central role that she gives to conscientious self-reflection it seems to us that the character of the evidence, independent of what the conscientious agent takes it be, drops out of the picture. In this way, then, her view is much more closely aligned with an egoist position. More on this below.

reluctant to adopt (i) or (ii) – she offers lengthy considerations against both egoism and egalitarianism – we suggest that her best alternative is to adopt (iii). We provide additional support for this suggestion below. This is no objection to Zagzebski's solution to the problem of disagreement. Rather, it is an invitation for Zagzebski to explain where her view lies on the spectrum between views that accord disagreement a great deal of epistemic weight in all cases, and those that don't. Of greater interest, however, is the substantive question, *How do we rationally resolve the problem of disagreement*? As we've already seen, Zagzebski's solution relies on the notion of conscientious self-reflective trust. In fact, this notion plays a prominent role throughout her book. Thus, in the remainder, we'll consider it in some detail.

DISAGREEMENT AND CONSCIENTIOUS SELF-REFLECTIVE TRUST

It will help to begin with Zagzebski's general picture of rationality. As self-conscious beings who desire truth, we are in the unfortunate predicament that there is no non-circular way to determine that our belief forming faculties and practices on the whole are reliable. Given this limitation, we must simply trust that truth is attainable and that our faculties and environment are suitable for its attainment. However, even given such trust, we sometimes recognize that our cognitive faculties are not providing stable or consistent deliverances. This conflict between our mental states will often – though not always – result in an experience of dissonance. And often our executive self will pre-reflectively attempt to resolve dissonance by removing or adjusting one or many of the items responsible for generating that experience. Zagzebski's notion of rationality is derived from this natural practice of resolving dissonance. For Zagzebski, rationality is just the 'property we have when we do what we do naturally, only we do a better job of it. To be rational is to do a better job of what we do in any case – what our faculties do naturally' (p. 30). Since we naturally resolve dissonance, our rationality is proportionate to how much we improve on this natural practice.

Not all dissonance is resolved automatically and without reflection. In addition to trusting our cognitive faculties, we trust that being deliberative and careful in using our cognitive faculties – that is, being conscientious – is the best way to satisfy our desire for truth. Our trust in reasons or evidence is ultimately derived from our more basic trust in

epistemic conscientiousness, since searching for evidence or reasons is something that a fully conscientious agent does. Zagzebski states:

> A conscientious person has evidence that she is more likely to get the truth when she is conscientious, but she trusts evidence in virtue of her trust in herself when she is conscientious, not conversely. Her trust in herself is more basic than her trust in evidence, and that includes evidence of reliability. The identification of evidence, the identification of the way to handle and evaluate evidence, and the resolution of conflicting evidence all depend upon the more basic property of epistemic conscientiousness. (p. 49)

When we experience dissonance and are unable to resolve it pre-reflectively, the rational response is to exercise epistemic conscientiousness in order to determine what is true. Interestingly, exercising epistemic conscientiousness does not just involve deliberate and careful consideration of *p*. For Zagzebski, the most rational response to dissonance involves consideration about the future and whether one's belief in *p* will survive 'future conscientious reflection, including reflection on future experiences, and future judgments about the past and present. The role of future conscientious self-reflection means that there is an important way in which the future justifies the present ...' (p. 50). In other words, one's belief that *p* is rational not because one has *judged p* to be true on the basis of conscientious self-reflection, but rather because one *now conscientiously judges that one will conscientiously judge p to be true*.

Disagreement produces dissonance within my deliberative reasons because though I trust my cognitive faculties and my conscientious use of my faculties, when I am conscientious I come to believe the following:

(1) *p*.
(2) Others have the same property (i.e. conscientiousness) that I trust in myself and believe not-*p*.
(3) This similarity to me gives me *prima facie* reason to trust them.
(4) Given my conscientiousness I have *prima facie* reason to trust myself.

The rational and conscientious agent, therefore, has to decide how to resolve dissonance by appeal to what she conscientiously believes will survive future conscientious self-reflection:

> Disagreement with people we conscientiously judge to be conscientious should be handled ... in a way that we conscientiously judge will survive conscientious self-reflection. ... In some cases the disagreement will eventually be settled by evidence both parties accept, but that is not the case for all disagreements, and in any case, what is relevant is not whether the disagreement will be settled by future or present evidence that somebody has somewhere. What is relevant for me is what I conscientiously believe, and what I predict will satisfy my future self-reflection, given what I conscientiously predict about myself. (p. 215)

This is further evidence that Zagzebski's view offers a kind of middle-way between the 'extreme positions' she describes. For on the face of it, in some cases I might conscientiously judge that my belief is most likely to survive future conscientious reflection if I trust my dissenter more than myself; in other cases, I might judge that trusting myself is the most likely to yield a belief that will survive future conscientious reflection; and in others still I might suspend judgment about this matter – and therefore suspend about the target proposition. This entails that neither egoism nor egalitarianism yields the correct result in all cases of disagreement. It also implies a result that defenders of the total evidence view have been eager to emphasize: there's no single correct solution to the problem of disagreement; that is, there's no single attitude (belief, disbelief, suspension of judgment) that is rational to hold in all cases of disagreement. Rather, what is rational to think will depend on the details of a given case.[5]

However, we are also now in a position to see how Zagzebski's view is *different* from the total evidence view and that her solution is extreme in its own way. The following close paraphrase reconstructs how she recommends dealing with the dissonance generated by the numbered propositions above:

> Awareness of this set of beliefs produces psychic dissonance and the desire to resolve the conflict by giving up one of the beliefs. [To] that end I will ask myself which beliefs are more likely to satisfy my future self-reflection. If upon reflection I trust (2) and (3) more than I trust (1) and (4), it is reasonable for me to adopt a skeptical attitude towards (1). ... But there are many other possibilities. I might trust (1) and (4) more than (2), in which case I would have no reason to give up (1) even

[5] See Thomas Kelly, 'Disagreement and Higher-Order Evidence', in Richard Feldman and Ted A. Warfield (eds.), *Disagreement* (New York: OUP, 2010) on this point.

> if I have a high degree of trust in (3). Alternatively, I might trust (1) and (4) more than (3). I have given reasons for adopting some form of the principle of trust in earlier chapters, but there is no reason to assume that my trust in that principle will exceed my trust in every other belief I possess. I have reasons for accepting (4) and may continue to accept it upon conscientious reflection, but maybe I can say the same thing about my belief (1).[6]

It is telling that in this passage the different possible rational responses are all within one case and within one subject. In other words, Zagzebski seems to not only allow that in some cases it is rational to retain one's initial belief while in other cases it is rational to suspend judgment, but also that for any particular case it can be rational to retain one's initial belief and it can be rational to abandon that belief. Rationality boils down to what I conscientiously trust will survive my future conscientious reflection, and there is no way of saying in advance where that trust will fall. In this way, then, her view seems much more closely aligned with an egoist position. The character of the evidence, independent of what the conscientious agent takes it to be, drops out of the picture in determining whether one is rationally handling a particular case of disagreement. On her view, however, what matters is not what I *have* conscientiously judged to be true prior to the disagreement but on what I *now* conscientiously judge to be most likely to survive my *future* conscientious reflection.[7]

CONSCIENTIOUS SELF-REFLECTIVE TRUST: GENERAL CONSIDERATIONS

What is distinctive about Zagzebski's view is the role she assigns to conscientious self-reflection, and more specifically to future conscientious self-reflection. On her view, such conscientious judgment about the deliverances of future self-reflection seems to be necessary and sufficient not only for rational belief in the face of disagreement, but also

[6] See pp. 216-17 for the exact quote. We have changed the numbering in this passage to reflect the numbered propositions above. Zagzebski gives a very specific example of disagreement involving eight propositions. For ease of exposition we have attempted to provide a general formula that captures the essence of Zagzebski's more specific example.

[7] The higher-level requirement here is explicit: 'Disagreement with people we conscientiously judge to be conscientious should be handled ... in a way that we conscientiously judge will survive conscientious self-reflection.' (p. 215)

for rational belief in any case.[8] However, there are reasons for doubting both of these claims, and additional reasons for doubting whether the appeal to our future selves is likely to provide the epistemic guidance we seek. But if Zagzebski's *general* views about the role of conscientious self-reflection are in doubt, then her views about disagreement are also insecure. Accordingly, we'll now consider the necessity, sufficiency, and helpfulness of Zagzebski's brand of self-reflection vis-a-vis rational belief.

Let's start with *necessity*. Is it necessary for rational belief that *p* that I now conscientiously judge that *p* would survive future conscientious self-reflection? Plausibly not. Consider a physicist working to confirm or

[8] In this section, we attribute to Zagzebski the view that conscientious reflection on one's future conscientious self-reflection is both necessary and sufficient for rational belief of the sort that interests her. We think that this is a fair interpretation of her view. However, we admit that it is not perfectly clear whether Zagzebski consistently means to invoke this higher-level requirement. In some passages, she seems to slide between first-level and higher-level requirements on rationality. For instance, on p. 50 she states: 'The line of reasoning of this chapter has the consequence that ultimately our only test that a belief is true is that it survives future conscientious reflection, including reflection on future experiences, and future judgments about the past and present.' Note that this sentence begins with the requirement that a belief *in fact* survive future conscientious reflection (a first-level claim), while the clause 'including reflection on future experiences' moves the discussion one level up. In light of other passages (e.g., the passage quoted above from p. 215, which is intended as an application of her general view to a specific case), we judge that on balance, Zagzebski favours the higher-level requirement we attribute to her. But because there is room for dispute here, we intend this section to accomplish the following *disjunctive* task: it either shows that Zagzebski's view is problematic on account of its higher-level requirement *or* that her view should be further clarified so as to explicitly reject such a higher-level requirement.

We hasten to add that merely renouncing the higher-level view may not solve all the problems that lurk nearby. Among these: suppose Zagzebski's official view is that rational beliefs must *in fact* survive future conscientious reflection, and that this is also sufficient for rationality. In that case, she will need to explain why surviving future conscientious reflection is necessary and sufficient (or even helpful) for obtaining rational beliefs, given that we're not often aware of what such reflection recommends. On the necessity point, it seems plausible that a belief can be made rational by a subject's evidence at time t_1 even if that belief would not survive conscientious self-reflection at t_2 (say, because the subject has new defeating evidence then). On sufficiency, it seems that even a belief that survives conscientious self-reflection at t_2 may not have been rational at t_1 (perhaps at t_1 the subject believes for bad reasons, but comes to believe upon better reasons later). Why think that reflection at t_2 justifies the subject's belief at t_1, especially if at t_1 the subject is not aware of the former? Finally, it is difficult to see how future conscientious self-reflection can provide present epistemic help, given that we are not often aware of the conscientious reflections our future selves will undertake. Thanks to an anonymous referee for pressing us to consider this 'first-level' interpretation of Zagzebski's view.

disconfirm the existence of the Higgs Boson. She designs her experiments in painstaking detail, and carries out the experiments with similar care. She repeats the experiments many times, and assesses the relevant data as conscientiously as anyone could. Finally, she forms the belief that the Higgs Boson exists. However, at no point in this process does she consider whether her belief in the particle will survive conscientious self-reflection on the part of her future self. Her belief seems perfectly rational, but Zagzebski's view seems incompatible with this judgment.

Worse still, Zagzebski's requirement on rationality appears to entail that few of our beliefs are rational. When reflecting on whether to believe some proposition, *p*, we usually only consider the first-order evidence indicating whether *p* is true or false. In our more careful and reflective moments we may also consider higher-order evidence regarding *p* (e.g. whether other reliable or unreliable people with regards to *p* believe *p*). But rarely – and for some of us probably never – when considering whether to believe *p* do we think about what our future conscientious self will also believe *p*. It follows, on Zagzebski's requirement, that few of us have very many rational beliefs.

Crucially, these problems arise not just when we compare Zagzebski's requirement to our everyday conception of rationality (though that would be problematic enough). Rather, the problems arise even on Zagzebski's comparatively more intellectualist view of rationality. On that conception, rationality consists in doing better what we do naturally. But as the above considerations show, we can do better – perhaps *much* better – than we do naturally, without making judgments about what conscientious judgments our future selves will make. For example:

- We might spend a great deal of time carefully and conscientiously evaluating whether our first-order evidence supports a scientific hypothesis;
- We might carefully and conscientiously consult both first and higher-order evidence in determining what to believe in a case of disagreement;
- We might devote an entire career to seeking and evaluating evidence relevant to God's existence.

Suppose that after such inquiries, we make judgments about the corresponding target propositions. On most evaluations, the resulting beliefs will be strong candidates for the post of *rational* belief. Of course, we sometimes make performance errors in evaluating our evidence,

so there's no guarantee of true or even rational belief at the end of our inquiries. But because the inquiries described above do not involve the very specific sort of self-reflection Zagzebski has in mind, the resulting beliefs are not even *candidates* for the accolade of rationality. No matter how carefully they are formed, if they don't result from conscientious reflection on the conscientious reflection of one's future self, these beliefs fail to be rational in Zagzebski's sense.

Now consider whether it is *sufficient* for rational belief in p that I currently conscientiously judge that I would judge p to be true on the basis of conscientious self-reflection. Here again, a negative verdict seems in order. Recall our scientist. Suppose that at time t1, in the midst of her research, she considers what her future self will believe at some later time t2 about the Higgs Boson, once this future self engages in conscientious self-reflection about the matter. At t1, she conscientiously judges that her future self will believe in the particle. She concludes that her future self will indeed hold this belief, and thus forms the belief herself. However, despite her conscientiousness at t1, she badly misjudges her evidence relevant to the claim that

> *Future:* My future self will believe in the particle after conscientious self-reflection.

Suppose that the weight of her current evidence actually supports the negation of *Future*, but that she ends up believing *Future* due to a cognitive malfunction that causes an egregious performance error. If this happens, then it seems her belief that *Future* is true is irrational, and consequently, that her belief about the Higgs Boson is irrational (at least if *Future* is its sole basis). Supposing this is right, we now have a case in which someone's current conscientious judgment about her future self's conscientious belief is insufficient to render that belief rational.

In addition to being neither necessary nor sufficient for rational belief, in many cases such higher-order reflection and future speculation is epistemically *unhelpful.* Consider again the scientist. She is hard at work in the lab and carefully evaluating the evidence for the existence of the Higgs Boson. It seems that turning her attention from the first-order evidence to what her future self is likely to believe upon future consideration is unhelpful with respect to her goal of believing the truth. If her colleagues catch her in the midst of such reflection, they will rightly exhort her to stop procrastinating and to get back to work. For surely what's most important in determining whether the particle

exists is the experimental data itself. If a judgment is to be made about the Boson, it should be made primarily on the basis of such evidence. If she indulges the distraction about her future escapades in conscientious self-reflection, our scientist merits a scolding. Reflection on her future self isn't making her a better scientist.

We can expand on this point by noting additional complications that may arise in the course of current conscientious reflection on the judgments of one's future self. First, it is unclear *which* future self should be the subject of one's reflection. Of course, it is absurd to suggest some uniform *temporal* criterion here. It is not as though what's relevant will always be reflection on what one's next-Tuesday-self or one's next-St.-Patrick's-Day-self will judge to be true. But if not these days, *when*?

We suggest that any plausible answer will feature not some specific temporal event, but rather, expectations about one's future *reliability*. Presumably, in consulting the conscientious reflections of one's future self, one aims to gain epistemic leverage from a source of information that is better than one's current self. This suggestion provides a way to clarify the epistemic relevance of reflection on one's future self.

But even here, problems lurk. Consider our scientist again. Either her future self (whichever one she happens to focus on) will be less reliable, equally reliable, or more reliable with respect to p than she is now. If she has reason to think her future conscientious self is *less reliable* with respect to p (e.g. perhaps she will be the unfortunate victim of some mentally disabling disease), then there is no good reason to now trust her future judgment, even if will be conscientiously formed, instead of her *current* judgment. If her future self is *equally reliable* regarding p, then appealing to her future self, at most provides a present-future case of peer agreement or peer disagreement. However, there is still no reason for her to defer to her future judgment. Rather, it seems to us that she should just reflect carefully on the current first-order and higher-order evidence (which will in this case involve evidence about her future self) regarding p and trust her current conscientious judgment regarding p. Therefore, only if she has good reason to think that her future self will be *more reliable* than her current self-regarding p would it be appropriate to defer to the authority of what her future self would conscientiously believe.

Even supposing she has good reasons to believe that her future self will be either equally or more reliable than she is now (not a small supposition since it seems that we are often without the requisite evidence to make

such judgments with sufficient reliability), often she will not be privy to what her future self will in fact believe with respect to *p*. For example, if her future self will be more reliable with regards to *p*, presumably this will be because either she will have evidence about *p* that she does not currently possess or because she will be better able to see some logical relation between her current evidence and p that she cannot currently see (or both). But since she does not currently have that evidence or that insight she will often be unable to judge what attitude her future self will take with respect to *p*. Therefore, the only way it could be rational to appeal to her future conscientious self when considering whether to believe some proposition is if 1) she has good reason to believe that she will be more reliable with respect to that proposition and 2) she has good reasons concerning what attitude her future self will have toward that proposition. Unfortunately, it is far from clear that these conditions are often satisfied.

Turning our attention back to the problem of disagreement, it seems to us that Zagzebski's solution is less than compelling for the reasons that her more general account of rationality is problematic. It seems to us that her almost exclusive focus on conscientious self-reflection comes at the cost of ignoring two important factors for rational belief: the character of our first-order and higher-order evidence, and how reliable we are in handling that evidence. Zagzebski's account seems incomplete to the extent that she ignores these items. For our part, we're confident that our future selves will have more information about the soundness of the above critiques once we have read Zagzebski's replies to them. In at least that respect, our future selves will be in a better epistemic position than our current selves. However, we are unsure what evidence Zagzebski will supply in responding to our arguments. Thus, reflecting on the conscientious reflection of our future selves is currently of little help to us – a fact which both gives us pause about the soundness of our criticisms and leaves us eagerly awaiting Zagzebski's replies.[9]

[9] Thanks to Tomás Bogardus and an anonymous referee for helpful comments and discussion.

EPISTEMIC AUTHORITY AND ITS CRITICS

LINDA ZAGZEBSKI

University of Oklahoma

The purpose of *Epistemic Authority*[1] is to defend the rationality of belief on authority from the modern assumption that the ultimate authority over the self is the self. More specifically, the strategy is to show that traditional epistemic authority, including the teaching authority of institutions such as the Catholic Church, can be justified even if we accept the modern value of autonomy.

The book begins with a brief historical and philosophical investigation of the rejection of epistemic authority, particularly as it is arises from the ideal of epistemic self-reliance. I argue that although that ideal dominates much of contemporary discourse, it cannot be defended from the work of Plato, Descartes, Locke, or Kant, and that the epistemological and moral arguments for self-reliance are weak. I then turn to an extended argument that the conscientiously self-reflective person is committed to authority in the realm of belief. I argue that epistemic (and emotional) self-trust is both rational and inescapable, that consistent epistemic self-trust commits us to trust in others, and that some of those others satisfy conditions for epistemic authority modelled on Joseph Raz's well-known theses of political authority. I apply epistemic authority to authority in communities, defend epistemic authority in the domains of morality and religion, and argue that the account of epistemic authority I give shows that epistemic authority is compatible with intellectual autonomy. Believing on authority is a demand of conscientious self-governance. It is not only compatible with autonomy, but follows from it.

[1] Linda Trinkaus Zagzebski, *Epistemic Authority: A Theory of Trust, Authority, and Autonomy in Belief* (Oxford: Oxford University Press, 2012). References to my book will be bracketed in the text.

EUROPEAN JOURNAL FOR PHILOSOPHY OF RELIGION 6/4 (WINTER 2014), PP. 169-187

In this essay I will give a summary of each book chapter, and will make brief comments on the critical essays most relevant to that chapter. For a few essays (Leszczynski, Anderson, and Benton) my comments are divided between two chapters. My main purpose is to give a guide to the essays in this issue, not to give a full response to each one. The authors raise a host of interesting questions on a wide range of topics pertaining to epistemic authority, religious authority, trust, and disagreement. I am grateful to the authors of these essays and to the editors of *EJPR* for this special issue. I hope that I will be able to have conversations at leisure with the authors, and that these essays will stimulate further work on epistemic authority.

CHAPTER 1. THE REJECTION OF EPISTEMIC AUTHORITY

In this chapter I trace the source of the rejection of epistemic authority to the modern rejection of authority in general, which has roots in events in the early modern era, as well as philosophical roots in the idea of autonomy. My particular interest in this chapter is the tendency to identify epistemic autonomy with epistemic self-reliance, which I find lacking in both historical accuracy and philosophical justification.

John **Cottingham**'s beautifully written, generous and insightful essay begins with some observations on this chapter. He agrees with me that a defence of epistemic self-reliance cannot be found in Descartes or Kant, and I thank him for pointing out that we get a strong rejection of epistemic self-reliance in Wittgenstein. Cottingham then raises some important questions about moral and religious authority. The radical historical contingencies of the development of our moral systems ought to make us sceptical of the use of our own conscientious reflection in judging the trustworthiness of moral and religious authorities, and he says that he thinks my project is incomplete unless it moves beyond epistemology to metaphysics and gives an account of moral truth. Otherwise, the authorities to whom we refer can easily have sensibilities that suffer from the same historical contingencies as our own. I agree that this is a serious danger, but in response, I would say that some of the wisest persons who have ever lived, including Jesus, Socrates, and the Buddha are still identifiable, and structures have been developed to protect their insights from the contingencies of each passing age. My account assumes that rationally self-reflective persons can recognize

wisdom even when it appears in different historical periods. Traditions that developed around these figures preserve wisdom, and operate as a counterbalance to the vagaries of cultural change. One of the functions of authority in a religious community is to protect the wisdom of the community from changes arising from historical accidents. Of course, as long as we are human, we are fallible, but I do not think we have an alternative that is any better than trusting the wisest among us, and our ability to identify them. Of course, we are on firmer ground if we can rely upon divine providence, but I was not appealing to divine providence for the audience of this book.

I am grateful to Damian **Leszczynski** for bringing a different philosophical background to bear on the issues of this book. According to my reading of his remarks, his views are much more congenial to my theses than he implies, and in any case, he probes some important issues of philosophical methodology and the interpretation of Descartes that are helpful to an understanding of the historical background on epistemic authority. I think Leszczynski's discussion of Descartes aids my theses in two ways. First, he agrees with my claim in Chapter One that Descartes is not a defender of self-reliance, and second, he argues that Descartes is a defender of self-trust, the importance of which I stress in Chapter Two. Descartes' grounds for requiring self-trust are not the same as the ones I use in the book, but it is valuable to be reminded of Descartes' argument of the need to be certain that God is not a deceiver. Both atheists and theists need self-trust, but for different reasons. Leszczynski says, 'I could have a real objective knowledge if I could correctly apply the method, but the method is reliable only when God is not a deceiver and when we were not created by accident.' Here again I think it is useful to consider the audience. What should we say to those people who believe we appeared through a purely naturalistic mechanism? I am arguing that even if we use a method that is neutral on that issue, we get a traditional conclusion about authority by unravelling the implications of reflective self-consciousness. I am afraid, however, that there might be a misunderstanding about the method I endorse. I do not mean to be appealing to common sense. However, I assume that we have access to the structure of rational consciousness, and that it is natural in the sense that we all have it. If the structure of human consciousness inevitably leads us to the objective source of our being, so much the better for my project.

CHAPTER 2. EPISTEMIC SELF-TRUST

In this chapter I begin with the reflective self governing itself in such a way that it has two aims: to make its psychic states fit their objects, and to make its psychic states fit each other. Both of these aims are natural. The first aim includes the aim to have true beliefs and fitting emotions. The second aim is to have a harmonious self, one without dissonance. The problem of epistemic circularity shows us that ultimately, the only test that we have succeeded in the first aim is that we have succeeded in the second. Since there is no guarantee that we have succeeded in the first aim without succeeding in the second, basic trust in our epistemic faculties is inescapable. Self-trust is rational because everything we call a reason for belief is derivative from what we do when we conscientiously reflect with the aims just given. Conscientious self-reflection is the basic norm of self-governance. Since self-reflection is natural, being rational is doing a better job of what we do naturally.

Pritchard and Ryan ask whether there is empirical evidence for my claims that we have natural desires, and that among those desires are a desire for truth and a desire for a harmonious self. They are also sceptical of the way I connect the natural and the normative. They propose that the way we naturally form beliefs and respond to conflicts within the self is best explained as part of an evolutionary story, and it would be odd if that coincided with the way we ought to behave. It is not part of my project to explain how evolution connects our cognitive faculties with the way the world is, but that is clearly a problem for others whose work I welcome. I do not think that what I am doing in this project awaits the outcome of work on the evolutionary theory of the mind.

Pritchard and Ryan then point out that I have not solved the problem of radical scepticism, but have shown that according to the view of rationality I propose, it is less rational to be a radical sceptic than to be self-trusting. I agree with them that that is what I do. I have no intention of resolving the sceptical problem, and I agree that what makes scepticism so worrisome is that it seems to be the consequence of what we do naturally– thoroughly seeking reasons for our beliefs. As I say in discussing Alston (p. 41), the person who desires full reflective justification for her beliefs and tries to attain it is doing what every reflective person does, only more thoroughly and scrupulously. When she does so, she finds that she is attempting the impossible. The issue for me is what response is most rational, and I give my reasons for thinking

the self-trusting stance is the most rational, but I have no solution to the paradox of scepticism, nor to Alston's paradox of the desire for full reflective justification.

Matthew Benton raises a related question in the section of his essay on rationality and the resolution of dissonance. I describe the situation of the reflective person who becomes aware of the problem of epistemic circularity. She trusts her faculties in her pre-reflective state, but when she realizes she cannot attain what Alston calls 'full reflective justification' for her beliefs, she must either achieve a higher level of trust, or she can live with the dissonance. I say that roughly, rationality is doing a better job of what we do anyway. Benton says he can see why trust after encountering the circularity problem is doing what we do anyway, but why isn't it doing a better job if we lose trust, given that we discover upon reflection that we lack something we rationally want? Alternatively, we could live with the dissonance since, as I point out, we can live with some dissonance. Not all dissonance has to be resolved.

I thank Benton for pushing me in the direction of living with dissonance. This is a point I wish I had made in the book. I do not want to say that a rational person must give up the desire for full reflective justification. She may not be able to do so. Often desires continue long past the time we realize they will never be fulfilled. But the issue for my project is what does she do about her pre-reflective trust? She must live in a psychic world in which it is impossible to satisfy a certain desire. If her psyche depends upon the satisfaction of that desire, then she will do what the radical sceptic (allegedly) does, and the structure collapses. But reflective trust permits her to continue managing her psyche as she always did, but in a more realistic way because now she knows that she has a desire for the impossible. But what if she does *nothing* in response to the awareness of circularity? Suppose that nothing changes in her pre-reflective trust in herself, and nothing changes in her beliefs or desires. She does nothing about it at all. She just doesn't think about it. I think that is the person Benton has in mind in one of his remarks. Can she live in a permanent state of dissonance without changing anything that generates the dissonance? It would be interesting to know whether such a thing can happen. I suspect something changes unconsciously. Either she gradually and with no conscious awareness becomes a sceptic, or she gradually and unconsciously realizes the critical function of self-trust in her psyche and it rises to the reflective level. But perhaps not. I don't think I can say what it is rational to do without knowing more about

what people actually do. This is an interesting possibility and I thank Benton for mentioning it.

Leszczynski also has some interesting comments on dissonance, and I like his use of Socrates to illustrate the advantages of experiencing dissonance in our psychic states. Because our consciousness is not insulated from the world around us, including the social world, we cannot get away with living the life of a coherent paranoid. We must continually adjust our conscious states – beliefs, emotions, etc. – in response to experiences of that world. I agree that dissonance is a good thing because it forces us to respond to it by changing something in our psychology– a belief, a putative memory, an interpretation of an experience, an emotion, and so on, and one of the reasons we need self-trust is that we think that in doing so, we are getting closer to having a mind whose states fit the world they are about– true beliefs, fitting emotions, veridical memories. I think that that is a good way to understand what it means to be rational.

CHAPTER 3. EPISTEMIC TRUST IN OTHERS

In this chapter I argue that consistent epistemic self-trust commits us to the same kind of trust in others. Two forms of epistemic egoism are incoherent, and we are committed to a weak form of epistemic universalism. The fact that another person has a certain belief always gives me a *prima facie* reason to believe it. If another person holds a belief conscientiously, I have a stronger *prima facie* reason to believe it. I distinguish two kinds of epistemic reasons: first person or deliberative reasons, and third person or theoretical reasons. Trust in self and trust in others are deliberative reasons for belief. These reasons do not aggregate with third person reasons, or what is often called evidence.

Charity Anderson's essay is delightfully clear and rewarding to read. In her discussion of this chapter she correctly observes that my universalist principle is weak. The fact that someone else has a belief p always gives me a *prima facie* reason to believe p, but the reason is not decisive and it can be defeated, perhaps even easily defeated. Nonetheless, it always has some weight. Anderson worries that the cases in which this principle has an effect on our beliefs are few, and that that reduces the effectiveness of the argument to oppose the egoist. She makes the parallel point about my argument that we owe a higher degree of trust

in the faculties of conscientious other persons whose conscientiousness we discover by being conscientious ourselves. Most people are not conscientious most of the time, and so, she argues, the principle does not commit us to very much.

I am neutral on the number of cases in which a universalist would believe what some other person believes and an egoist would not. But even when the universalist and the egoist coincide in believing or not believing some other person, I think that there is a world of difference between the person who approaches the epistemic faculties and beliefs of others with the attitude 'Innocent until proven guilty', and someone who approaches other persons with the attitude 'Demonstrate your innocence', or even worse, 'Guilty until proven innocent'. The first is like the attitude of the epistemic universalist, the second is like that of the standard epistemic egoist, and the third is like that of the extreme epistemic egoist. I think that universalism not only can make an important difference in certain critical cases of belief, but it forces us to become intellectually humble. A disposition to see others as basically equal to ourselves, and hence partners in the search for truth, can mitigate our natural tendency to inflate the self, and to close our mind and heart to the riches of other minds and the results of their reflective efforts to get the truth. Anderson is right to mention that self-deception can lead us to misjudge our own degree of conscientiousness, and I think that the trust in others to which our self-reflection leads can also have the effect of making us more realistic about our own defects in the use of our powers.

In Chapter Two I argue that ultimately, our only test that a belief is true is that it survives future conscientious reflection. That is because the problem of epistemic circularity means that we never have a final verdict that a belief is true, and there is always a possible gap between our evidence or reasons for belief at any given time and the truth. But we think that if a belief is true, it will not be disconfirmed by future evidence. Since I argue that many different kinds of psychic states can function as deliberative reasons for a belief, ultimately, the only way I have to tell that a belief is true is its survival of my conscientious reflection now and into the future when my experiences change, and I have given up some of my other beliefs. So the future confirms or disconfirms what I believe now. To believe p is to think p is true, and to think p is true is to make a bet on our future reasons for belief. I think we make these bets all the time. If what I believe now is disconfirmed by future evidence, then I will have to change my mind, and there is nothing wrong with

that. Change is a necessary part of the process of making our beliefs closer and closer to the truth. But sometimes I can be confident that the future will not disconfirm my present beliefs. This is particularly true for beliefs that are central to the self, such as moral and religious beliefs, and beliefs closely tied to one's personal commitments. I think that this is important for the resolution of disagreement with other persons. I do not offer a formula for resolving disagreement, but stress the fact that it is a dilemma that arises within the consciousness of each individual conscientious person. Anderson says that we will not always know the conscientious way to respond, but that is as it should be if it is resolved by a self-directing person. Granted, it is hard to predict what will survive our future reflection because we do not know what future experience will bring, but to some extent we decide our future self. Each person has to decide where she will hold the line – 'This belief is part of me' – and where she will leave it open that the belief may have to be given up.

CHAPTER 4. TRUST IN EMOTIONS

This chapter gives arguments for trust in emotions parallel to the arguments of the previous two chapters. We have the same kind of reason for basic trust in our emotion dispositions as we have for basic trust in our belief-forming faculties, and we have the same kind of commitment to trust in the emotions of others. This makes emotion disagreement a problem parallel to the well-known problem of epistemic disagreement. Trust in admiration gives us another route to trust in the *beliefs* of others. Intellectual admiration can be a deliberative reason to trust the deliverances of the epistemic faculties of admirable persons.

CHAPTER 5. TRUST AND EPISTEMIC AUTHORITY

This chapter moves from trust in the beliefs of others based on their similarity to ourselves to trust in the beliefs of epistemic superiors, persons that we conscientiously judge are more likely to get the truth than we are. This leads to a defence of epistemic authority modelled on Joseph Raz's theses of practical authority. The two main theses are the Pre-emption thesis and the Normal Justification thesis. The *Pre-emption thesis* is a thesis about what it means to take someone as an epistemic authority. According to that thesis, the fact that the authority has a belief p

is a reason for me to believe p that replaces my other reasons relevant to believing p and is not simply added to them. Raz's *Normal Justification thesis* says, roughly, that the normal way to show that A is an authority for S is to show that S is more likely to act on his first order reasons if he does what A says to do rather than to try to act on those reasons directly. I have two *Justification of epistemic authority theses* that are analogues of Raz's NJ thesis. The first says that the authority of another person's belief for me is justified by my conscientious judgment that I am more likely to form a true belief and avoid a false belief if I believe what the authority believes than if I try to figure out what to believe myself. The second says that the authority of another person's belief for me is justified by my conscientious judgment that I am more likely to form a belief that survives my conscientious reflection if I believe what the authority believes than if I try to figure out what to believe myself. I argue that even the epistemic egoist should accept the first thesis. When someone else satisfies one of these theses, I should take her belief pre-emptively.

I was delighted to read **Arnon Keren**'s paper on epistemic authority and pre-emption since he has already engaged with these issues. Keren agrees with my central claims, but argues that there are important differences between practical and epistemic authority that damage the line I take in defending the justification of epistemic authority. Keren agrees that Raz's pre-emption thesis applies to epistemic authority, but he argues that some of Raz's other theses of authority do not. Keren appeals to our practices of criticism to support the claim he and I share that epistemic authority is the power to generate pre-emptive reasons for belief, but he strengthens the position by claiming that epistemic authority is the normative power to make it a *duty* for others to believe pre-emptively. I prefer to refer to what a rational, self-governing person 'should' do, and do not use the language of duty, but my reasons for not speaking of duty have to do with my views about duty in general that I do not bring up in the book, and I see no reason to object to Keren's position. If there is any distance between us on this point, I do not think it is serious.

Keren does differ from me on Raz's No Difference thesis. According to that thesis, there is nothing that subjects ought to do as the result of the exercise of authority which they did not have reason to do anyway. The authority's directive merely gives them new reasons for doing what they already had reason to do. Raz argues that the No Difference thesis is false for authority in general. Keren argues that the thesis is false for

practical authority but true for epistemic authority. I argue that the No Difference thesis is false for epistemic authority as well as for practical authority. When a practical authority tells me to do X, I have a reason to do X that is not simply another reason to do what I already had reason to do. Similarly, when an epistemic authority tells me that p, that gives me a reason to believe p that is not simply another reason to believe what I already had reason to believe. I do not understand Keren's claim that the No Difference thesis applies to epistemic authority because he seems to be highlighting a different, but very interesting difference between practical and epistemic authority. In the case of epistemic authority, Keren points out, the authority cannot give me a reason to believe p unless she has one herself. In contrast, a practical authority can give me a reason to do X when sufficient reasons to do X did not previously exist. Keren is right that this is an interesting difference that deserves investigation because it applies to the current controversy about the way knowledge is transferred via testimony. It is a difference in the conditions for the exercise of responsible authority, a difference that applies to the person in authority, not to the subject. However, I do not see that it pertains to the No Difference thesis.

Keren agrees with me that epistemic authority gives me pre-emptive reasons to believe what the authority tells me, but Keren disagrees with my defence of pre-emption by referring to Raz's financial shares argument in which Raz argues that unless one pre-empts in taking the authority's advice, one's track record will be worse overall. Keren points out that in the practical case there are only two options: sell the stock or not sell the stock, whereas in the epistemic case, there are three options: believe p, believe not p, or suspend belief. As long as our epistemic goals include both getting truth and avoiding falsehood, he argues, then withholding judgment, while not the best outcome, allows us to guarantee an epistemic outcome that is second-best. So instead of pre-empting (always believing the expert), we are better off in withholding judgment in at least some of the cases in which the expert and I disagree. So Keren's idea is that I have to weigh three possible outcomes: (i) believing a truth, which is the best outcome, (ii) believing a falsehood, which is the worst outcome, and (iii) withholding belief, which means I might lose the chance for the best outcome, but at least I avoid the worst.

I think this is an excellent case for discussion of pre-emption because there are a number of different scenarios that deserve discussion. I think Keren is right that the subject has three options, and I thank him for

pointing that out, but notice that the authority has three options also. So the authority can say 'Believe p', 'Believe not p', or 'It's inconclusive whether p, so withhold judgment'. It will still turn out that my track record is better if I follow the authority pre-emptively under these conditions. That is, I should believe when she says to believe, and withhold when she says to withhold. But there are other scenarios that raise William James' famous point in 'The Will to Believe' about weighing the value of truth against the disvalue of falsehood. For instance, if my hatred of falsehood exceeds my love of truth, it would be reasonable for me to withhold belief more often than I would otherwise. We must, then, be clear about what our first order epistemic ends are.

Benjamin McMyler also pursues the topic of pre-emption in his essay, arguing persuasively that authority includes more than the power to give pre-emptive reasons, but also includes an essential interpersonal dimension. McMyler begins with a worry that the authority of someone else's belief is not strong enough to be parallel to Raz's view of practical authority. The reason is that Raz says that authority is exercised in the giving of commands or directives, which means that the parallel exercise of epistemic authority is the authority of testimony, which I discuss in Chapter Six, rather than the authority of belief, which I discuss in Chapter Five. I think that our disagreement here is minimal. I am willing to say that the authority of belief is weaker than the authority of testimony, which has the addition of important interpersonal features, but there is something interesting about Raz's Normal Justification thesis even if authority is exercised in the intentional issuing of directives. If someone knows better than I how to act on my first order reasons, she knew it before she told me anything. So her satisfaction of the Normal Justification thesis does not require that she testify that p to me, and that is why someone's belief can be authoritative for me, and it is also the reason that inanimate objects like GPS systems can be authoritative for me. Nonetheless, I agree with McMyler that the interpersonal dimension adds something important to authority, something that explains why it is stronger than the authority of navigation systems, and why it has often raised worries about the compatibility of authority and autonomy. When I obey someone who is authoritative for me, I am submitting to that person. Trust between authority and subject is essential to the proper operation of authority. Paradoxically, authority is an aspect of conscientious self-governance because I need others who intentionally and effectively help me govern myself according to standards I accept.

The conclusion is that McMyler is right that the normative power to give others pre-emptive reasons is not sufficient for authority in its most interesting sense.

Trent Dougherty defends evidentialism and the Lockean view that faith is belief on the evidence of the testimony of an expert. He says that the view that faith is a kind of belief on a kind of evidence is 'run of the mill'. He sees no reason why a subject of authority would take the authority's testimony pre-emptively when she already has reasons that bear on the issue. He does not comment on the Razian reason that we are more successful at reaching our own ends if we pre-empt. However, he says he doubts that it is even possible to pre-empt, and I agree that the psychology of pre-emption deserves close investigation. If pre-emption is essential to obeying authority, and if it is impossible to pre-empt, it would follow that it is impossible to obey authority. If pre-emption is not essential to obeying authority, and if the subject merely considers the authority's directive as another piece of evidence, it is hard to see how obedience is anything other than doing what a rational person would do in any case. Dougherty says he freely made a vow to obey the Church when he was confirmed, but I would like to discuss with him his reasons for making the vow since it seems to me likely that he judged that the Church satisfies some thesis of the justification of authority similar to the ones I propose. (In fact, his reasons were probably more reflective than mine were since he was confirmed as an adult, whereas I was only twelve).

The main dispute between Dougherty and me, however, is on the nature of evidence, and what it means to believe on evidence. He says that evidence is anything, broadly construed, that aims at truth. As I mentioned above, I argue in Chapter Two that since we have no guarantee that anything we call evidence (or reasons for belief) leads us to truth, evidence is less basic than self-trust in our epistemic faculties. Nothing we call a reason or evidence would be a reason unless it is reasonable to trust the connection between the conscientious use of our faculties and getting the truth. Self-trust *is* a reason, and because it is a reason, all the other reasons we can identify have the status of being reasons, and that includes everything Dougherty calls evidence. The difference is that self-trust reveals the first person states that we take to indicate truth in addition to propositional evidence. Emotions can be in that category, and that is why I say that trust in the emotion of admiration can give us another route to the justification of authority. But even if all we are

doing is weighing evidence in the Lockean sense, we get an argument for the reasonableness of taking certain persons as authoritative on the Justification theses. What I think Dougherty adds is the need to distinguish degrees of authority, an important thing to do when there are competing authorities, or when I am not sure whether someone else is so much more likely than I am to get the truth in some domain, that I should consistently, and for the foreseeable future, take their testimony in that domain pre-emptively. But I think that even though the idea of degrees of authority is an important issue, it overlooks something important about authority – the interpersonal dimension stressed by McMyler. For some authorities, I invest myself in them. I do not say, for instance, I am 70% sure of you, and only 68% sure of some other religion. As Coady argues below, it is hard to say that the authority of the Church is all of nothing, but it can't be a matter of dividing a certain percent of my loyalties between one putative authority and another either.

In the second half of **Anderson**'s paper, she brings up some interesting questions about pre-emption. One question that I find particularly fascinating is the question of counterfactual stability. Have I acted on authority if I let the authority's directive pre-empt my other reasons, but I might very easily not have done so? I had not thought of this question before and have no immediate answer. If counterfactual stability is not necessary, that seems to mean that we can act on authority accidentally. But if counterfactual stability is necessary, the conditions for acting on authority become very strong, perhaps too strong. I hope that other writers on pre-emption have something to say about this question.

CHAPTER 6. THE AUTHORITY OF TESTIMONY

This chapter gives a parallel justification for the authority of testimony, which differs in some interesting ways from the authority of someone else's belief. I defend a view of testimony that is strongly anti-reductionist, similar to the views of Moran and McMyler.

In the second section of Matthew Benton's paper, he brings up a kind of 'authority' that is possessed by a testifier who knows what she tells me. I do not address knowledge or the transfer of knowledge in my book, and I do not treat the testifier who knows p as possessing authority in the sense I mean. Benton is right that the conditions for successful testimony he discusses are third personal, and the pre-emption thesis does not apply. But in this chapter I am interested in a subset of testimonial cases

that satisfy the justification thesis, and which pose some issues that are an extension of the issues discussed for the authority of someone else's belief in the previous chapter. The problem for most of the book is from the first-person perspective, although I argue for a third-person version briefly. The problem of this chapter therefore has very little in common with most of the literature on testimony.

Benton's discussion of testimony brings up an interesting issue in the last section of his paper, in which he discusses testimony as evidence vs. the trust model of testimony as applied to religious belief. Readers will find a very nice discussion of how we can know that a revelation is from God in the first place, prior to the operation of the trust model.

CHAPTER 7. EPISTEMIC AUTHORITY IN COMMUNITIES

Authority is most interesting when it involves networks of people in communities with a structure of authority that serves the purposes of the community. This chapter begins with differences between political authority and authority in small communities with a high degree of trust. These differences permit the justification of a stronger kind of authority than we get from Raz's Justification of Authority principle, and can apply to epistemic authority.

The epistemic authority of my community is justified for me by my conscientious judgment that if I believe what We believe, the result will survive my conscientious self-reflection better than if I try to figure out what to believe in a way that is independent of Us.

Tony Coady's paper is relevant to this chapter, as well as to Chapter 9, and I discuss it below.

CHAPTER 8. MORAL AUTHORITY

In this chapter I argue that the justification of epistemic authority applies to moral beliefs, whether the authority is an individual or a community. Epistemic egoism is very common as applied to moral beliefs, and even extreme egoism has a substantial number of adherents. In fact, it is probably the dominant view. I argue in this chapter that moral epistemic egoism is incoherent. Taking a moral belief on authority is justified in the same way taking non-moral beliefs on authority is justified. Moral understanding, in contrast, cannot be attained through the testimony of authorities, but even understanding can be aided by other persons.

Cottingham's paper, discussed under Chapter One above is relevant to this chapter, and Coady's paper, discussed under Chapter Nine, is also relevant to the topic of moral authority.

CHAPTER 9. RELIGIOUS AUTHORITY

This chapter argues against religious epistemic egoism, and uses the defence of religious epistemic universalism as a way to generate a novel form of the *consensus gentium* argument for theism. I discuss divine testimony and give three models of revelation: the model of chains of testimony about an original experience of the divine, the model of recurring first-person experience of the divine, and the 'high point' model in which a high point of history after an original revelation creates a tradition. I conclude with a justification of the religious authority thesis which says that the authority of my religious community is justified by my conscientious judgment that if I engage in the community, following its practical directives and believing its teachings, the result will survive my conscientious reflection better than if I try to figure out what to do and believe in a way that is independent of Us.

Three essays address issues raised in Chap Nine. John Schwenkler and Jacek Wojtysiak discuss models of revelation. C. A. J. Coady discusses authority in religious communities. Coady's comments are also relevant to Chapters Seven and Eight.

John Schwenkler discusses two problems I raise for the first model of revelation. On the chain model, nobody is as justified in a belief acquired through a chain mechanism than the first person in the chain. The farther away we are from the source, the less justified we are. Schwenkler quite rightly points out that actual religious traditions always add many features to this model that increase the justifiability of the belief: there are multiple chains, there is back-tracking and double-checking, convergence on a single person from multiple directions, and most importantly, supervision of the entire process. These features bring the model much closer to the third model I describe, in which there is a development of the tradition over time under the supervision of a teaching authority. But the feature Schwenkler adds that is closest to the high point model is what he calls Summation: Each original witness is given only a proper part of the message, which is put together when the chains converge. The third model I describe is similar to Schwenkler's up to the point of convergence. On my third model, what is transmitted is not an original

experience, but an interpreted body of oral and written traditions put together from various sources at a high point that is thereafter transmitted to succeeding generations. The transmission is intended to have continuing relevance to future members of the community, and with an authoritative method to preserve the high point of the tradition. James Kugel argues that Jewish tradition should be understood on this model, and I argued that Catholic tradition is also. As Schwenkler points out, the chain model does not explain how the transmission of tradition is improved, yet we ought to accept its improvement from other sources of truth in addition to sacred texts and ancient theological documents.

In another essay on models of revelation, **Jacek Wojtysiak** plausibly argues that the experience model is a refinement and expansion of the chain model, and the high point model is also dependent upon the chain model since the community's beliefs need to include a reliable story on the initiating event of divine intervention. Furthermore, the concept of a high point implies a structure in which the chain is modified to a chain of transmission of a way of understanding God. This must involve both recalling remote events, and applying the community's ancient exemplars of faith and practice to every believer's life.

Tony **Coady** has written a very sensitive discussion of communal authority, particularly in the way it can become corrupted when institutionalized. If I understand him, he does not object to my theses of epistemic authority as applied to communities, but argues that I apply them in a way that seems to lead to an 'all or nothing' acceptance of an authority, which is unrealistic for an institution like the Catholic Church. My model starts with the authority of an individual, then moves to authority in small communities, and I then apply the model to the authority of the Church and other large religious communities that have become institutionalized. Groups of people can recognize the same authorities, and practical and epistemic authority can often become intertwined. Somebody can be better than I am at both determining how to reach my practical goals and how to get true beliefs, and there can be general agreement about this among a group of people. Rich and varied communities can arise in this way. But when the authority gives a large number of practical directives as well as testimony to a large number of beliefs, complications are bound to arise, as Coady rightly points out. The community might split apart for what seems to us now to be a trivial matter (such as the East/West dispute over the *filioque* clause in the Nicene Creed), yet it might not split apart over more serious matters

of sexual morality, contraception, and abortion. How can that happen? My view is that there is both personal and communal self-reflection, and they do not necessarily coincide in their results. I may conscientiously come to the conclusion that some particular teaching of my Church does not satisfy my conscientious reflection upon my total set of beliefs, values, and experiences, and so the Church is not more likely to reach the truth about that particular matter than I am myself. But of course, if I make such a judgment very often, that will weaken my belief that in general the Church is more like to reach the truth (or get a belief that will survive my conscientious reflection) than I am myself. So Church members may disagree about certain doctrines and still remain members, but one of the things they may disagree about is what they need to agree about in order to remain members. This is a problem in ecclesiology. Every community has to decide what makes the community what it is, and what authority structure preserves it as what it is. The Church has a communal set of beliefs, values, historical memories, and goals upon which the community continually reflects, and its authority structure is one of the things upon which it reflects, sometimes leading to important changes, as happened during Vatican II.

Coady rightly notes that I do not discuss institutional authority, and he brings up some important problems that arise once a political dimension is introduced. He suggests that that can lead to a separation between the wisdom dimension of authority and the institutional dimension, where the persons with power in the institutional structure are not the wisest (or holiest) persons. This is an important point, and I thank Coady for pressing me on it. I think that institutional structure is itself justified by communal reflection, but when that means that those in power lead the reflection, there is a strong bias in favour of maintaining their power. However, there are always voices of wisdom, not the least of which is the Pope (and it is interesting that Pope Francis has arranged his life in such a way that he cannot be 'protected' by those around him from finding out what he needs to know, a particular problem for the previous two popes). I also think it is important that a community, whether institutionalized or not, must not succumb to what I call communal epistemic egoism (Chapter 10, sec. 4), the analogue of personal epistemic egoism, and that gives a community the obligation to conscientiously listen to and respond to criticism from the outside as well as dissent from within. I think it is also worth noting that the ideal authority structure is not necessarily democratic since democratic institutions are subject to the

same kinds of corruption lamented by Coady, and in any case, epistemic authority by its nature cannot be democratic. I realize that many people are mistrustful of institutions in general, but I think that institutions are necessary for the perpetuation of moral and religious insights. Coady is obviously right that power corrupts, but power serves some important human purposes as well.

CHAPTER 10. TRUST AND DISAGREEMENT

This chapter addresses the difficult problem of how to handle conscientious disagreement with other persons whom I conscientiously trust. Given that trust in others is a commitment of the conscientious self, and given that I conscientiously judge that some others are equal or superior to myself in their ways of getting the truth, there can be a conflict between a belief I conscientiously hold and a conscientious belief held by someone with an opposing belief who is at least as trustworthy as myself. Going back to the distinction between theoretical and deliberative reasons, I argue that the problem is not mysterious if it reduces to a conflict of theoretical reasons since there are many cases in which evidence points in conflicting directions. The real problem is a conflict among our deliberative reasons, which are irreducibly first personal. The problem can only be resolved by determining which is most conscientiously trusted. The last section of the chapter turns to communal epistemic egoism and the need to resolve disagreement between communities in a way parallel to the resolution of conflict within the self.

At the beginning of the chapter I reject two extreme positions on this issue – the egoist position that the beliefs of others do not count, and the egalitarian position that my reasons for belief count no more than the reasons of others. But I also reject compromise positions which on many issues can lack the theoretical support of either of the extreme positions. I call the disagreement problem an antinomy because I think it shares some features of Kant's antinomies, where the solution is not for each side to give a little to the other side, but to change the terms of the debate. The outcome may be that one side 'wins' in a sense, but not on the terms of the original dispute. **Joshue Orozco and Nathan King** give a very helpful summary of the 'total evidence' position of Thomas Kelly and others, which is a well-motivated middle position that is not a mere compromise between extremes, and as they observe, is not far from my own view. But since the problem in its strongest form arises from my

first-person viewpoint, Orozco and King think that I am closer to the egoist position because 'evidence drops out' and is replaced by survival of conscientious reflection.

At this point in their essay I think that Orozco and King have a misunderstanding of my view of rationality as survival of future conscientious reflection. The norm of reflection is not to believe what my future self will believe, but to believe in a way that survives my conscientious reflection on into the indefinite future. As I said in discussing Chapter Two and Anderson's essay above (Chapter Three), anything we call reasons for belief derive from what we do when we conscientiously reflect with the aim of truth, but since there is never a final verdict on whether we have been successful in reaching the truth, we aim to believe in a way that survives reflection on into the future when we may have more evidence. Orozco and King describe a physicist attempting to confirm or disconfirm the existence of the Higgs Boson. She conscientiously reflects upon the data she has and forms a judgment that it exists. Orozco and King say she is rational in her belief and I agree, with one provision. If she has reason to believe that the data will differ in the future, she should form her judgment provisionally. But of course, she may not know whether or not the data will change, in which case her belief that the Higgs Boson exists includes a bet that her belief will not be disconfirmed by future evidence. As I say above, I think we make these bets all the time. There is nothing mysterious about them. But one thing she does not do is to think about what her future self will think. I agree with Orozco and King that she is not doing that. Furthermore, I say in the book that it is not even necessary that she ask herself what will survive her future conscientious reflection. It is necessary that she forms her beliefs in a way that will survive reflection, but it is not necessary to ask herself, 'Will this belief survive conscientious reflection?' She need not raise the second-order question about her own level of rationality very often.

CHAPTER 11. AUTONOMY

In the final chapter of the book I argue that conscientious self-reflection is the fundamental norm of autonomy. Attacks on self-trust are attacks on autonomy. The conscientiously self-reflective person described in this book is the autonomous agent. Since conscientious self-reflection shows us the rational justification of belief on authority, authority is compatible with autonomy and, in fact, authority is entailed by autonomy.

BOOK REVIEWS AND NOTICES

MARK MANOLOPOULOS
Monash University

Paolo Diego Bubbio & Philip Andrew Quadrio (Eds.). *The Relationship of Philosophy to Religion Today.* Cambridge Scholars Publishing, 2011.

This collection of essays from a variety of leading and promising thinkers has several things going for it. Stylistically, the texts are clearly written and thus refreshingly accessible. Secondly, the editors have included contributions not only from both the analytic and Continental traditions, but also from theistic, atheistic, and agnostic perspectives, so the volume exemplifies an openness to a variety of currents of thought – an inclusivity that we should expect/demand *today*. And perhaps most importantly, the essays are rigorously argued, engaging, and life-relevant, so that what we have is, in fact, quite an expansive exploration into not just 'one' relationship between philosophy and religion today but into a number of *relationships*. And so, the book actually lives up to the promise of its title (and perhaps surpasses it in some ways).

To begin with, the most surprising and impressive part of the book for me is the form and content of the editors' Preface. Right from the start, and with a nice stylistic mix of humility and ambition, the editors challenge us thinkers of religion to be more ambitious ourselves, by not just limiting ourselves to traditional philosophico-theological problems (divine non-/existence, the problem of evil, etc.) but of asking (at the risk of 'arrogance'): 'What *ought* the relationship between philosophy and religion be?' (p. vii). The editors are very clear on the directions they want philosophy of religion to pursue today (and tomorrow): they speak in particular of the 'political' and 'socio-political' several times throughout the rest of the Preface, as well as referring to 'the symbolic

and regulative dimensions of religious life, the existential and cultural import of religion, and the question of religion and politics' (p. ix) – expansive, indeed. (The question of the relationship between religion and the political shall recur throughout this review.)

I will say a thing or two to say about all the papers, but will have more to say about the contributions that I myself find most relevant, particularly in terms of the most essential and urgent relationships that philosophy is beginning to have – and should have – with religion. The first essay is exemplary in this regard. Matheson Russell's 'Philosophy of Religion in a Secular Age: Some Programmatic Reflections' begins by offering a concise overview of the four basic directions or categories of philosophical thinking of religion: metaphysical, epistemological, philosophico-theological, and philosophico-anthropological. Russell offers a nuanced critique, one with which I am in agreement; i.e. that philosophy of religion has 'become increasingly abstract and technical' (p. 13). The overview alone is impressive, but then the author goes on to situate the various strata in relation to their broader socio-intellectual contexts, with an emphasis on their relationships to secularity. Confirming the thoughts and aspirations of the editors, Russell insists upon the need that philosophy of religion consider its relation to its 'political, social and cultural dimensions': this phrase (and its variations) is repeated throughout the paper. And I was particularly encouraged to note that he cites one particular (and crucial) aspect of this contextualization: 'the adoption of capitalist modes of production, and the development of concomitant forms of socialization and individuation' (p. 12). (Inspired and informed by the likes of Slavoz Žižek, Antonio Negri, Alain Badiou, and others, thinkers of religion are today beginning to critique capitalism.) The next piece is John Bishop's 'Philosophy and Religious Commitment'. This is a solid piece, which is no surprise, given that the author is a well-established figure in contemporary philosophy of religion. In this essay, Bishop begins by convincingly dismantling elements of Alvin Plantinga's 'Reformed epistemology', before outlining a 'modest fideism' influenced by William James and developed by Bishop.

The third contribution is Paul Crittenden's 'Faith In Keeping With Reason: A Critique of the Regensburg Address'. The papal Address (delivered in 2006), which appears to be a strong affirmation of the relationship between rationality and divinity, is critiqued from the outset and along various fronts; e.g. the Pope's attempt to portray a strong link

between biblical faith and reason; his construction of a wider gap between Islamic and Christian thought than what may be the case; etc. Crittenden thus effectively weakens the papal argument in a way that is both rigorous and enjoyable. But perhaps what I found most engaging about the piece is what I perceive as an absolutely critical task for both philosophy in general and for philosophy of religion today (and tomorrow): of the need for a revised/expanded figuration of reason, one that avoids, one the one hand, a narrow yet bloated scientistic-instrumentalistic hyper-rationalism, and, on the other hand, an impotent reason diluted by a host of excesses (hyper-relativism, over-contextualism, an excessive emphasis on difference and otherness, etc.), thus denying reason its force and universality. Crittenden cites the likes of Edmund Husserl, Ludwig Wittgenstein, and Jürgen Habermas as thinkers contributing to the reconception of reason (p. 70).

The next contribution is Kevin Hart's 'Contemplation: Beyond and Behind'. This is a typically brilliant piece of work from Hart, with all the hallmarks: a careful retracing of a concept over the centuries, an expansive/encyclopedic knowledge of the subject-matter, and of course, beautiful prose. But just as this essay is the most beautiful piece in the collection, it is also perhaps the most abstract/removed when it comes to its relation to the rest of the volume – though Hart's text certainly has a lot to say about the continuing relationship of contemplation to theology and philosophy.

The fifth essay is Graham Oppy's '"New Atheism" versus "Christian Nationalism"': this text also exemplifies the talents of its author: clearly written, thoroughly researched, rigorously argued with all the necessary provisions, qualifications, and nuances . . . in sum, a 'no-nonsense' approach to thinking religion – which should surely be one of the defining characteristics of philosophy of religion today. This piece explores the 'New Atheist' attack on religion, and it is refreshing – even heartwarming – to observe an atheist with a fierce intellect undermining the excessive claims of the New Atheists. Of course, one may find objections with the essay – and Michael Levine certainly does. 'New Atheism, Old Atheism and the Rationality of Religious Belief' is quite a *tour de force*, somewhat reminiscent of Nietzsche (which is a good thing), but risking a condescension which should have no place in philosophy or philosophy of religion today; e.g. Levine construes philosophy of religion's relation to 'mainstream philosophy' as 'a quaint

and poor relation' or 'an irrelevant anachronism' (p. 157). As to who is (more) 'right' would require an extended response, but one way in which I would summarize this most engaging debate is that Oppy may be too forgiving, whilst Levine may be too severe. (I would also contend that philosophy of religion today should re-cast this particular debate in the following way: New Atheism is dogmatic, which makes it nothing new, whilst religion is guilty of some of the charges made by New Atheism, and must be re-figured or even re-made as an open, minimalist – and, yes, rational – faith, which is/would be something new.)

The seventh chapter is 'Religious Reasons in Political Debate: Jeffrey Stout and the Tradition of Democracy' by Anthony J. Langlois. Taking up the theme of religion and politics called upon by the book's editors and Russell, Langlois explores this relationship in the context of liberal democracy. As the chapter title indicates, he outlines and evaluates the work of Jeffrey Stout, who wrote the landmark work, *Democracy and Tradition* (2004). (I myself find it increasingly difficult to defend liberal democracy – particularly in its capitalist manifestation – in the wake of ecological, financial, and a multitude of other crises.) The essay is another solid piece of scholarship. The final essay is Douglas Pratt's 'Religious Identity and the Denial of Alterity: Plurality and the Problem of Exclusivim'. The question of religious diversity should certainly be considered when exploring the relationship/s of philosophy to religion today, so this is a welcomed contribution. But what stands out about this text on this topic is that it offers a nuanced understanding of exclusivism: that it should be perceived as being located on a 'continuum' with its 'competing' categories of inclusivism and pluralism (p. 202), and of distinguishing between subtler/more sophisticated forms of exclusivism from exclusivistic extremism (p. 203). Once again, this essay is characterized by the essential features of good scholarship and reflection.

Of course, *The Relationship of Philosophy to Religion Today* is not – nor does it pretend to be – an exhaustive exploration into all of the actual or possible relationships. But it is certainly encouraging to note that the work signals some of the most relevant (and interesting) directions. (As for myself, two particular directions resonate most sharply: the need for a re-figured reason, one that is simultaneously humble and ambitious and informed by a certain openness to faith; and the articulation of how this faith and reason may contribute to much-needed socio-political change.) I strongly recommend this book for anyone interested in philosophy of religion.

AKU VISALA
University of Notre Dame, IN.

Justin Barrett, *Born Believers: The Science of Children's Religious Belief*, Free Press, 2012

In the last decade or so, new literature on the biology and psychology of religion has accumulated. Some popular treatments of this literature have already been published: Pascal Boyer's *Religion Explained* (2001), Justin Barrett's *Why Would Anyone Believe in God*? (2004) and Robert McCauley's *Why Religion is Natural and Science is Not* (2011), to name just a few. *Born Believers* (henceforth, BB) by psychologist Justin Barrett is a new addition to this literature. In the book, Barrett, while bracketing the evolutionary issues and some of the cognitive ones as well (such as ritual and practice), focuses on the development of religious belief in childhood. His argument is, simply, that children are in fact naturally prone to adopt religious beliefs. In addition, he wants to fend off suspicions that his conclusion might produce in the minds of religious sceptics and maintains that religious belief is not 'childish'.

To be sure, BB is targeted at a popular audience, so the reader should not expect detailed descriptions of research methods or too much theoretical background or philosophical reflection. The tone of the book is light and the text is peppered with amusing (and sometimes less amusing) stories. However, for those of us who are not too keen to trudge through great amounts of psychological research articles and reports, Barrett's short summaries of various experiments are useful and there is enough depth to the description of studies. Actually, one of the most interesting aspects of the book are the numerous descriptions of the ingenious experimental designs that developmental psychologists have devised to understand the beliefs and mental lives of young children and even babies.

The main thesis of BB is that 'children are prone to believe in supernatural beings such as spirits, ghosts, angels, devils, and gods during the first four years of life due to ordinary cognitive development in ordinary human environments' (p. 3). Barrett also makes the somewhat more controversial suggestion that children might have a bias towards more specifically Judeo-Christian-type god-concepts. Indeed, 'evidence exists that children find especially natural the idea of a nonhuman creator of the natural world, possessing superpower, super knowledge, and super perception, and being immortal and morally good' (p. 3).

In order to understand BB's basic argument better, it is useful to briefly examine what Barrett means by 'natural' in this context. Often naturalness is associated with something being innate or hard-wired, but this is not what Barrett means by it. Instead what is being referred to are actions characterized by cognitive ease and automaticity. Robert McCauley has made a useful distinction between practiced naturalness and maturational naturalness. Maturationally natural capacities emerge normally in human development and are mostly independent of specific cultural influence. Learning to walk and speak one's native language are the most typical examples of maturational naturalness. Whereas maturationally natural capacities require no special instruction, training, artefacts or tools, there are capacities that require exactly these. McCauley holds that capacities like reading, effortlessly driving a car or doing basic algebra are natural in the latter (practiced) sense: they require specialist instruction, training and tools.

Given these distinctions, BB's argument is that certain kinds of religious beliefs are maturationally natural to children, that is, children tend to adopt certain kinds of beliefs given that they are reared in a normal environment and have normal biological makeup. This does not result into any kind of biological determinism, because maturationally natural capacities are not necessarily determined by our biology. As Barrett puts it, 'just because we have a biological disposition towards a trait does not mean it will develop without the right kind of environment, and just because something is not built in does not mean that it is not nearly inevitable as a part of human development' (p. 19).

Now, BB consists of two parts. In the first part, Barrett goes through various developmental studies to support his argument, and in the second he considers some of the implications of his argument.

He first reviews evidence that suggests children finding and positing agents easy and natural. With early-developing capacities, like agency detection and theory of mind, thinking about invisible agents and their special powers is cognitively easy for children, which in turn makes it likely that they will adopt god-concepts, if they are around. He then describes several studies (some of his own) on how god-concepts that include superknowledge and immortality might be natural for children to understand.

Barrett pits some of his conclusions against the received view of the religious development of children, that of the developmentalist Jean Piaget. According to Piaget, children anthropomorphize God, that

is, they think about God as if God were just like another human (like their parents in most ways). Only when capacities for abstract reasoning develop around eight or nine years of age do they begin to distinguish between God and humans. Barrett argues that Piaget got something the wrong way around. It is not the case that children learn to think about God by extrapolating on the basis of their parents, but instead children before the age of four tend to think that all agents are like God, super knowing, immortal and super powerful. So the more sophisticated children's reasoning gets, the more children learn to restrict their intuitions and understand, for example, that not all people know what they themselves know and that biological life has an end point. So, instead of anthropomorphizing God and gods, children might actually make their parents and other agents god-like.

One of the main points in the first part is that the common assumption of children simply soaking up whatever their parents teach them is wrong. Against this (what Barrett calls the) Indoctrination hypothesis, Barrett defends the Preparedness hypothesis, namely, that children have a natural cognitive tendency to adopt religious-type concepts and not just any religious-type concepts but specific kinds of god-concepts. In other words, children do not learn everything that their parents teach them with similar ease, but instead some ideas make more sense to them than others.

For instance, Barrett argues that because teleological and design thinking is so intuitive for children (and adults too), teaching children evolution is rather hard. Children have the tendency to assume several things that make understanding evolution difficult. First of all, they assume that there is inherent teleology in nature. This is shown by series of ingenuous experiments conducted by Deborah Kelemen, among others, that suggests children (and adults) prefer functional explanations for natural kinds and substances. Second, children also tend to think that there is a link between agents and order: order does not come about all by itself but is a product of intentional agency. Third, children also have a strong intuition that natural kinds and entities are not human made: children can readily distinguish between artefacts (human made) and natural entities (designed but not human made). These tendencies give rise to a kind of 'intelligent design' bias, that is, a tendency to see nature and especially animals as designed for some purpose. Having this tendency makes it rather counterintuitive to believe that the natural world is a product of biological evolution, which is not driven by any

agent or does not operate with a goal in mind. Thus, even in cases where children have grown up in an environment where evolution is explicitly taught and supernatural belief actively discouraged, children have trouble understanding and believing evolution and favour agent and designer type explanation over standard biological ones.

The second part of the book deals with some of the implications of the naturalness of religious belief hypothesis. Barrett begins the part by rarefying the main point: not all god-beliefs are maturationally natural, but only certain basic ones. The core 'natural religion' includes the belief in some sort of creating and super powerful god and design in nature but it does not, for example, entail strict monotheism. Furthermore, at least some of the attributes of orthodox monotheistic gods are not supported by maturationally natural cognition, but are, in fact, contrary to it. For example, the Trinitarian view of God and the notion of God being outside time and space are rather unnatural and, thus, more difficult for both adults and children to believe. Here we encounter some interesting studies about what Barrett calls the 'theological incorrectness effect'. Some theologically correct attributes of God, like omnipresence and 'omniattention' seem difficult for people to process in situations where reflection is not especially encouraged. Several studies revealed that although religious believers verbally profess that God is everywhere and can do all things at the same time, they nevertheless made systematic distortions to stories about God's actions and presence. They tended to 'anthropomorphize' God in ways that made God more like normal agents, such as being able to direct attention to only to one place instead of some other or being in one place rather than another. The theological incorrectness effect, thus, reveals the difference between natural religion and theologies. Natural religion, according to Barrett, provides an anchor-point or the basic building blocks for religious elaboration and actual theologies.

Some critics of religion, such as Richard Dawkins, have famously argued that teaching religion to children is a form of child abuse. Given what Barrett has said so far, it is not difficult to figure out that he disagrees. Although he acknowledges that religious education can be in some extreme cases abusive (as all education), there is nothing abusive in teaching children what their parents genuinely believe and practice. This applies to both religion and non-religion. Barrett then makes a pragmatic argument for this. Because children have a natural tendency to seek role models and wanting to become like their parents, if parents

do not teach their children what they themselves believe and profess, the children will naturally feel excluded. Behind the worry of Dawkins and others is the view that children are blank slates and learn anything that their parents teach them. But if the research that BB presents is correct, this assumption is false. The reason why most children adopt the religious beliefs of their parents is not that their parents and their religious communities indoctrinate, bully or abuse them to believe, but simply because it is easy for children to adopt religious ideas, if they are around.

Finally, I want to briefly discuss one philosophically salient topic in BB. Given BB's basic claim, the natural question to ask is whether the fact that we do indeed have some maturationally natural capacities that make the emergence of certain kinds of god-beliefs likely provide problems for the truth or justification of religious beliefs. More specifically, the philosopher of religion is interested in whether psychological explanations of religious beliefs will have any significant impact in the philosophical debates about atheism and theism.

This question is pressing, because naturalistic explanations of religion are often used as parts of critiques of religion. It also looks at least intuitively plausible that if we can explain the emergence of a belief in such a way that the explanation makes no reference to the truth of the belief, the belief itself is undermined. In other words, if we can explain how god-beliefs come about without making references to gods or such, god-beliefs must be false or unjustified. In the first case, the argument would be that the fact of naturalistic causal explanation for religious belief gives us a reason to think that religious beliefs are false. The problem, however, is that such an argument would commit the genetic fallacy. There is no logical contradiction between the statements 'religion is completely naturally caused' and 'God exists', so we should not infer straightforwardly from a naturalistic explanation of religion to the falsity of religious belief. As to the justification of religious beliefs, things are more complicated. It seems at least intuitively plausible to think that if I learn that the causes of my god-belief is purely natural, this would constitute a defeater for my belief and I would not be any longer justified in believing in God.

Apart from assuring that the BB account does not by itself debunk religious belief, Barrett does not really deal with these issues. There are a few hints towards an answer to the justification problem, though. Barrett points out that many trustworthy beliefs that adults have

(e.g., permanence of solid objects, other people have minds, gravity) are based on maturationally natural cognition. From this a general rule is derived: 'our minds as basically trustworthy to deliver true beliefs and ... our naturally arising "childish" beliefs should be regarded as true until we have good reason to suspect them as being problematic' (p. 173). More philosophically put, deliverances of our cognitive faculties in their proper conditions are prima facie justified and since religious belief is a product of such faculties, it should be considered as prima facie justified. Here we have a kind of Reidian (after Thomas Reid) answer to the justification problem.

There are some worries here, though. First, might there be something in the BB account itself which functions as defeaters for the prima facie justification. One could, for instance, argue that a benevolent and all-powerful God would be unlikely to use such fallible mechanisms as hair-triggered agency detection to provide people with beliefs about God. There might be some divine hiddenness-type worries about the goodness and trustworthiness of God. Nevertheless, some (including Barrett himself in his other works) have argued that the Christian God would most likely create natural psychological mechanisms that would favour at least some kind of belief in God. Reformed theologians often talk of *sensus divinitatis* as fulfilling such a role. If this were plausible, the BB account of religious beliefs (if true) would be (albeit modest) evidence for the existence of God.

The second problem is that prima facie justification of the kind suggested by BB is rather weak, especially in a cultural context where alternative religious beliefs to one's own (and even a wide variety of non-religious worldviews) are readily available to all. So, without any independent reasons for one's religious belief other than the basic, intuitive support of one's own cognitive mechanisms, specific religious beliefs would be unlikely to survive for long. From the remarks that Barrett makes elsewhere in BB, we can infer that he believes that some such reasons could, nevertheless, be given and at least some types of religious beliefs might not be any worse off than most non-religious beliefs. However, the original point still stands: even if the BB account does not all by itself make religious beliefs unjustified, epistemically responsible religious (and non-religious) believers need to reflectively consider defeaters in order to maintain justified belief.

In conclusion, let me just say that a book that is as well written and concise as BB deserves a readership outside psychological circles. From

a philosophical point of view, it does not offer much, especially in terms of the problems relevant for philosophy of religion. But again, this might be too much to expect from a popular book such as BB. Fortunately, there is an emerging literature on philosophers engaging with psychological and biological explanations of religion (e.g., *The Believing Primate*, OUP 2009). Further, Barrett himself has addressed the topic in several articles and his other 2012 book *Cognitive Science, Religion and Theology: From Human Minds to Divine Minds* (JTF Press), which I would suggest as a companion piece to BB.

ULRICH SCHMIDT
Munich School of Philosophy

Charles Taliaferro. *Dialogues about God*. Rowman and Littlefield Publishers, 2009.

In his excellent book *Dialogues about God*, Charles Taliaferro gives a comprehensive introduction into the main questions concerning the subject of God. The theist Taliaferro presents his introduction in the form of a dialogue, because he appreciates the abilities of self-questioning and of placing oneself into the opposing position (p. xii). He chose the form of a friendly dialogue in order to enable a constructive discussion and reduce the hostility which sometimes occurs in philosophical discussions (pp. xiii-xiv). The four characters of this dialogue are the secular naturalist Pat, the theist Chris, the agnostic Tony and the negative theologian Liz who holds that God is beyond human concepts (pp. xiv-xvi).

Is Theism coherent and valid as an Explanation?: Pat begins the dialogue by arguing that theism is incoherent. Theism assumes the existence of God, a conscious immaterial person. But we are only familiar with bodily beings. Without a face there can be no grin. Without a body there can be no thinking, feeling and acting. The idea of an immaterial person is incoherent (pp. 2-4). Chris objects to that and argues that human beings are conscious immaterial persons. Hence conscious immaterial persons are possible. Chris argues that materialism with respect to human beings is false. Beliefs, purposes and desires cannot be reduced to physical processes in the brain. We can imagine that human beings are zombies, i.e. that they behave as if they had mental life with conscious experiences while in fact lacking mental life altogether. This conceivability is

evidence for the assumption that zombies are possible. If zombies are possible, then my body has the property of possibly existing without my consciousness, whereas my consciousness does not have this property. Thus, by Leibniz's Law of the Indiscernability of Identical Entities my body and my consciousness are different entities, because they differ in at least one property. We know that there are many correlations between our bodily states and our mental states. However, correlations are not the same as identity relations (pp. 4-5). Pat responds to this by arguing that the constant correlations between our bodily states and our mental states are evidence of the identity of bodily and mental states (p. 5). Chris admits that in the case of water and H_2O the constant correlation of the two is evidence of the identity of the two, because we can grasp the identity by thinking that water is composed of H_2O. By contrast, we will never observe subjective conscious experience. We can only discover it introspectively. Therefore we know that our conscious experiences are radically distinct from our brain states (p. 7). Chris further argues that even if 'human consciousness is necessarily physical, it does not follow that every form of consciousness is necessarily physical' (p. 8). Even if materialism with respect to human beings were true, it would not follow that conscious immaterial persons are impossible.

Then Pat and Chris discuss whether theism is a valid explanation of the cosmos. Chris argues that the natural sciences explain the existence and the properties of many contingent objects in the cosmos. But we also need an explanation for why there is a cosmos at all and why there is this cosmos rather than another cosmos. For if every part of the cosmos is contingent then the cosmos itself is contingent (p. 13). Theism is an explanation of our cosmos. The existence of the cosmos is explained as created according to God's contingent will. God himself does not have to be explained, because he exists necessarily (p. 14). Pat's stance is that the necessary truths (including mathematical truths) are features of language and that we could simply define the cosmos as existing necessarily (p. 15).

Eventually Liz presents her view that God is not a thing among other things. He is beyond all human concepts and categories. Unlike Pat, Liz does not deny God's existence. Liz only denies that God can be described (p. 19). Chris responds that the negations Liz proposes only make sense if there is some positive concept of God (p. 21).

The Concept of God: The divine attributes are discussed in the second conversation. God is omniscient and has immediate awareness of all

states of affairs (p. 27). Pat argues that God's omniscience is incompatible with human libertarian freedom by presenting the following dilemma. Either God is omniscient or not. If he is, then he knows your future actions. That you will go sailing tomorrow is fixed now, because God foreknows it. Hence you do not have the ability to do otherwise than going sailing tomorrow. You lack freedom in the libertarian sense. If God is omniscient, then human beings do not have libertarian freedom. Hence, either human beings do not have libertarian freedom or God is not omniscient (p. 29). Chris's response is that being omniscient means to know all truths that are possible to know. Today there is not yet a matter of fact about what you will do tomorrow. Hence it cannot be known and is not required to be known in order to be omniscient (p. 30). This stance seems problematic for the theist, however. For there are many prophecies in the Bible foretelling future events. For example, Jesus foretold Peter's denial and the prophets of the Old Testament foretold Jesus' life, suffering and death. There seems to be better solution for the theist, namely Molinism. Luis de Molina assumed that before the creation of the world God knew about every possible creature what she/he/it would do if put in certain circumstances. God chose to create this cosmos on the basis of this Middle Knowledge. God's foreknowledge of what the creatures in this cosmos will do does not determine their actions, because it has no causal influence on the creatures. Rather, God's foreknowledge is dependent on what human beings will freely choose to do. (See Thomas Flint, *Divine Providence: The Molinist Account* (Ithaca: Cornell University Press, 1998), pp. 36-46.)

Pat takes the discussion to the divine attributes of omnipotence and essential goodness. She argues that these attributes are incompatible. For if God is essentially good then because of his nature God cannot do any evil. But Pat can do evil and thus is more powerful than God (at least in this respect). Hence God is not omnipotent (p. 36). Chris responds that the ability to do evil is not a power, but a deficiency (p. 37).

Arguments for the Existence of God: In the third conversation Chris advances four arguments in favour of theism. First, Chris presents the ontological argument. If God exists, then he exists necessarily. Essential or necessary existence is one of the qualities of a perfect being. God cannot exist contingently. Therefore either he exists necessarily or his existence is impossible. The existence of God is possible. We have reason to think so, because there is no contradiction in the concept of God. This is so for two reasons. First, this is so, because we can positively

conceive that God exists. Second, this is so, because the complexity of the universe, the existence of human consciousness and the existence of religious experience are evidence in favour of God's existence and thus also in favour of the assumption that it is possible that God exists. Hence the existence of God is not impossible. Thus God exists necessarily (pp. 52-53). Pat objects that Chris simply defines God into existence. Existence and necessary existence are not proper qualities of a concept. For they do nothing to enlarge the concept (p. 54). In Chris's view, however, necessary existence is a proper quality of a concept, because it is an excellence which contingent beings lack (p. 54).

Second, Chris presents the cosmological argument. Pick any physical object in our cosmos you like. Let us assume we seek an account for the existence of this physical object. It exists but it could also not have existed. It exists contingently. It does not explain its own existence. It has been caused by another physical object which again exists contingently. This physical object again exists contingently as well as its cause, etc. We can go back as far as we wish in this kind of causal chains of contingently existing objects. But we never get an ultimate explanation of the existence of any of these contingently existing objects. There are only contingent objects in the cosmos and the cosmos as a whole exists contingently. Only God as necessary being provides an ultimate explanation of the existence of the cosmos and every object in it (pp. 56-57). In response to this Pat keeps her view to regard the cosmos as the ultimate frame of explanation. She is not willing to go beyond physical explanations to explain the cosmos (p. 62).

Third, Chris presents the teleological argument. The cosmos has a high degree of structure and order. This enables the existence of irreducibly complex organisms with consciousness. The existence of an intelligent good God provides an intentional explanation for this. It is likely that God wants there to be intelligent conscious creatures in the cosmos and thus designs to cosmos accordingly. By contrast, the naturalist has a hard time of explaining the structure, order and complexity of the cosmos, because only causal explanations are available for him (pp. 63-65). Pat objects that this argument is only plausible if we posit a finite amount of time. If there is an infinite period of time, then all possible configurations of the cosmos are expected to occur sometimes (p. 65).

Fourth, Chris presents the argument from religious experience. There is widespread testimony of the existence of a transcendent reality. This is evidence in favour of the existence of a transcendent reality. Pat raises

four objections and Chris responds to each of them. First, Pat argues that because of their diversity religious experiences cannot be evidence in favour of one and the same entity. Chris responds that the religious experiences of adherents of different religions share many elements. They have in common that God reveals himself as good, compassionate and powerful (p. 68). Second, Pat argues that the existence of religious experiences can be explained naturally without appeal to the existence of a transcendent reality. Religious experiences can be explained by the human desire for mercy, forgiveness, etc., and by the religious training the persons with religious experiences have received. Chris responds even if God uses natural means to perpetuate religious experiences it does not follow that religious experiences do not point to the existence of God (p. 70). Third, Pat argues that even if religious experiences are evidence in favour of a transcendent reality they are not evidence in favour of the theistic God as a necessary, omniscient, omnipotent, and essentially good being over other transcendental realities. Chris agrees with this objection by Pat and admits that it is not by experience alone but only by reasoning and revelation by God that we come to know the divine attributes (p. 71). Fourth, Pat argues that religious experiences are not valid evidence, because we can produce them by pharmaceuticals. Chris responds that even if this is so it does not undermine religious experiences as evidence for the existence of God. God is omnipresent and religious experiences can be evidence of him even if we have the power to produce them (p. 72).

The Problem of Evil: Pat starts the fourth conversation by presenting the problem of evil. If God is all-good he wants to prevent evil. If God is all-powerful and all-knowing he can prevent evil. So why is there evil? (p. 77) Chris answers this question with the free-will defence. It is good that God creates creatures with free will. Free will entails that the creatures have the ability to do evil. God allows that in order to keep the value of free will for his creatures: ' ... deep moral freedom involves acting in favour of goodness when one could do otherwise.' (p. 81)

The second challenge Pat raises is about the amount of evil. Why did God not create a world with less evil? Chris argues that there is no best possible world. This is analogous to there being no greatest possible number. One can always ask why there is not less evil and why there are not more values in the world (p. 80).

Miscellaneous Topics: In the fifth conversation miscellaneous topics about God are discussed. Miracles are events brought about by special

acts by God that violate at least one law of nature. Miracles like the resurrection of Jesus are radical breaks of the regularities of the world (p. 106). Pat distrusts testimonies of miracles in general, for she thinks it is more probable that the witnesses had hallucinations than that there was a radical break of a law of nature (pp. 107-108). Chris regards the special divine acts involved in miracles as similar to acts of human free will. Both types of actions are free and cannot be predicted by knowledge of the current state of the universe and the laws of nature. The type of miracle that occurs most often is religious experience in which God causes an experiential awareness of the divine in human beings (p. 108). Chris's view that miracles are possible can be backed up by Alvin Plantinga's argumentation. First, miracles are not incompatible with classical Newtonian science, because the laws of classical mechanics hold only for closed systems. (See Alvin Plantinga, *Where the Conflict Really Lies: Science, Religion and Naturalism* (Oxford: Oxford University Press, 2011), pp. 76-84.) Second, miracles are compatible with quantum mechanics for the same reason. Furthermore, some physicists think that miracles are not even breaks of the laws of quantum mechanics, because the wave function of quantum mechanics only gives us probabilities where each particle is located. On this view miracles are extremely improbable events caused by God. (See Alvin Plantinga, *Where the Conflict Really Lies: Science, Religion and Naturalism*, pp. 94-97.)

Eventually, the incarnation as an issue within Christian theism is discussed. Jesus Christ is both God and a man. The divine attributes are necessary existence, omniscience, omnipotence, omnipresence, essential goodness and eternal or everlasting existence. By contrast, the human attributes are contingent existence, finite knowledge, power and presence, not essential goodness and temporally finite existence. These properties are direct contraries. Pat argues that it is impossible for one person to exemplify both the divine and the human properties (p. 111). Chris responds by stating that the Christian doctrine is that Jesus Christ is both wholly God and a whole human being. But he is not the whole God and he is not a mere human being. Chris advocates Eastern Christology. There is one single divine nature. But God as Father, Son and Holy Spirit has three centres of consciousness. In the incarnation the Son kept all the divine attributes. But he undertook a radical self-limitation and limited himself to a human mind within his divine mind (p. 112). The Son of Good took on an individual human nature. In a proximate sense it was Christ's human nature which did miracles and rose from the dead while

ultimately it was the Son of God who did miracles and rose from the dead. (See Thomas Flint, '"A Death He Freely Accepted": Molinist Reflections on the Incarnation', *Faith and Philosophy*, 18: 1 (2001), 5-6.) Here is how this account can be applied to the attribute of essential goodness. Jesus Christ as the Son of God was essentially good and not able to sin. But Jesus as a human being was not essentially good. He was able to sin and faced real temptation. By his Middle Knowledge of counterfactuals of creaturely freedom God foreknew that Jesus Christ as a human being would resist all temptations if he was put it the circumstances he was put in. In this way God foreknew with certainty that Jesus would resist all temptations and freely choose to accept the death on the cross for the sake of the redemption of us sinners. (See Thomas Flint, '"A Death He Freely Accepted": Molinist Reflections on the Incarnation', pp. 7-10.)

Conclusion: Charles Taliaferro's book *Dialogues about God* is an excellent, comprehensive and easy-to-read introduction into the important topics about God. Taliaferro presents the different views with clarity and covers the most important aspects of the topic. It is a great introduction for undergraduate students and non-philosophers, and gives a wonderful overview of the fascinating topic of God. It is not intended go into great depth and detail and it is not intended to advance the current discussions in philosophical theology.

GRAHAM WOOD
University of Tasmania

Fraser Watts (ed.). *Creation: Law and Probability*. Ashgate, 2008.

Creation: Law and Probability is a collection of papers drawn from, or prompted by, the second meeting of the International Society for Science and Religion held in Boston in 2004, and published within the Ashgate Science and Religion Series.

With a keen interest in the relation between science and religion and particular interests in the nature of physical law (probabilistic or otherwise) and the nature of chance (physical or otherwise), I found this book interesting. It gave me valuable insight into how various religious perspectives understand the concepts of law and probability, and the role those concepts play within those perspectives. (Although it should be

noted that monotheism is the dominant perspective represented.) I was particularly interested in how law and probability were understood with respect to themes such as freedom, fruitfulness, openness, and purpose.

However, while various religious conceptualizations of probability and laws of nature were well presented and examined in detail, I was disappointed by the lack of detail with which the secular conceptualizations of probability and laws of nature were presented and examined. In general (there were exceptions), the various contributors to the volume made a simple two-way distinction between laws understood as related to necessity, or related to regularity. But this two-way distinction does not capture the detail of the contemporary debate concerning laws of nature. For example, contemporary philosophy of science would identify at least a three-way distinction among theoretical positions when analyzing the metaphysics of laws of nature: dispositional essentialism, nomic necessity and regularity theory. For example, see Alexander Bird's *Nature's Metaphysics: Laws and Properties* (Oxford: Oxford University Press, 2007). And given that a number of contributors, for example Clayton (pp. 39-41), had themselves used a three-way distinction when considering the relation between laws (understood from a theological perspective) and the god of monotheism, I felt that there were interesting parallels that could have been examined more explicitly. And, unfortunately, the detail of the discussion with respect to probability was also disappointing. Contemporary philosophy of science distinguishes up to five interpretations of the probability calculus: classical, logical, frequency, propensity, and subjectivist. For example, see Salmon et al, *Introduction to Philosophy of Science* (Indianapolis: Hackett, 1999). And, I would like to have read about the compatibility of the various philosophical interpretations of probability with respect to (say) the actions of a monotheistic god.

Put simply, given the book's intended purpose (as noted in the preface), of enriching the dialogue between theology and science, there was an unfortunate imbalance in the detail of the content. I found the theological content interesting and engaging, but I suspect that a philosopher of science would find the lack of engagement with contemporary theories about laws of nature or interpretations of the probability calculus frustrating. If the detail with which some of the theological themes were examined had been matched by similar detail with respect to the examination of the secular understanding of natural laws and probability this book would have been a more significant work.

A person reading this book with a good understanding of contemporary philosophical theories about laws of nature or the interpretations of probability could apply that understanding to the theological analysis within the book itself. But the book would have been much stronger with the addition of a chapter on laws of nature (and how they relate to scientific explanation), and a chapter on interpretations of probability, both written from the perspective of contemporary philosophy of science.

None-the-less, the book is a valuable contribution to the dialogue between science and religion and provides insight into how our scientific understanding of laws and probability might be accommodated within a religious worldview.

Introducing the book, Fraser Watts identifies law and probability in *nature* as central concepts within a secular discourse, and purpose and freedom in *creation* as central concepts within a theological discourse. The interaction of law and probability is presented as a way of understanding how nature can fulfil a divine purpose. Watts also offers an outline of the chapters in the book, dividing them into three groups. Chapters 1-3 are described as philosophical, Chapters 4-7 as scientific and, Chapters 8-10 (together with the Afterword) as theological.

In Chapter Two, Peter Harrison provides a fascinating insight into the historical development of the concept of laws of nature, and highlights the important role the early modern concept of laws of nature still plays in our understanding. Beginning with the Aristotelian view that considered mathematics and science as distinct enterprises, Harrison describes how the role of mathematics, previously understood as simply an instrument for calculation, changed in scientific practice. As nature came to be seen as a machine, mathematics became the language with which it was described. Harrison ends by drawing the reader's attention to the, perhaps underappreciated, theological commitments behind our commonsense understanding of laws of nature.

In Chapter Three, Philip Clayton presents an optimistic view of the possibility of 'broad explanatory consonance' between religious belief and scientific study. Clayton identifies three major theological positions on laws of nature, namely laws as: eternal necessities; necessities imposed upon the world by divine choice; and patterns that humans detect in the natural order (pp. 39-41). The chapter explores theological interpretations of law in a number of detailed and interesting ways, but does not match that analysis with equally detailed exploration

of contemporary interpretations of law discussed by philosophers of science. As mentioned above I would like to have seen what Clayton characterizes as the three-way theological distinction considered in parallel with the three-way distinction among dispositional essentialism, nomic necessity, and regularity theory.

Chapter Four, by George Ellis, examines the so-called fine-tuning of the universe and the postulation of a multiverse as an explanatory response to the apparent fine-tuning. He considers the possibility space of universes and asks: what determines the range of possibilities, and what determines the universe (or universes) that is (are) instantiated from within this possibility space? Importantly Ellis calls our attention to the status of any answers to these questions. Are they scientific answers, or are they some sort of meta-scientific answers? After all, scientific explanations are often built out of scientific laws, but (surely) the scientific laws of this universe cannot explain the coming into being of those very laws. Ellis uses the phrase 'meta-cause' to point to this meta-scientific explanation. The chapter concludes with a discussion of the significance of the possible existence of a multiverse with respect to a number of theological and philosophical themes.

In Chapter Five, Niels Gregerson offers an interesting exploration of self-organization in and of the natural world. He suggests, rather than relying solely upon universal laws to understand the world, we should add general formative principles and causal capacities. Doing so, he claims, would allow for 'a fertile avenue for theological explanation of a self-developing world' (p. 82). Furthermore, Gregerson discourages the view that self-organization is anti-religious and points to the role that God might have 'in, with, and under' natural processes (p. 91). He suggests that divine action should be thought of not as a triggering activity but a structuring cause.

Chapter Six, written by Michael Ruse, is largely an examination of the concepts of teleology and progress in evolutionary thinking. Ruse examines the concept of progress in evolution with reference to a number of thinkers. Richard Dawkins is quoted as claiming that evolutionary progress occurs when new evolutionary possibilities come into being via major innovations such as 'the origin of the chromosome, of the bounded cell, of organized meiosis, diploidy and sex' (p. 113). In contrast, Steven Jay Gould is reported as resisting the notion that evolution is progressive, preferring the notion of 'directionality' that 'comes about through random processes and nature's constraints. Life

is a bit open ended. It started simple. It cannot get less simple. It can get more complex' (p. 119).

In Chapter Seven, Nancey Murphy addresses the free-will problem. She suggests that the free-will problem is either misunderstood or badly formulated. She sees reduction as the problematic presupposition in both compatibilist and libertarian responses. She advocates replacing reductionism with downward causation (or whole-part constraint), where lower level (either deterministic or indeterministic) processes generate variation that is then selected among by upper level processes. She claims that while most organisms have this structure, additional cognitive capacities in humans allow for free-will. These additional capacities, of self-transcendence and reason, are dependent on language and are built upon capacities that are goal directed and evaluative.

In Chapter Eight, David Bartholomew examines six topics in which probability arguments are relevant to issues of theological significance: the origin of life; God's existence; the fine-tuning of the universe; Dembski's explanatory filter; God's action in the world; and the operation of chance being within the providence of God. Unfortunately this chapter is weakened by the lack of engagement with the detail of the various interpretations of the probabilities being used in each of these arguments. None-the-less Bartholomew's discussion is of real interest. For example, I found it genuinely illuminating to learn Bartholomew's opinion (when considering the possibility that God might act 'rarely' within the bounds of a probabilistic law such that the law was not violated by God's action) that a theist would find such behaviour 'un-Godlike'. I also found interesting Bartholomew's examination of the suggestion that God might use chance within a creative process.

In Chapter Nine, Wesley Wildman advocates 'a new kind of natural theology, one that is comparative in approach and prizes transparent criteria for the sake of correcting and guiding a dynamic process of inquiry' (p. 177). Wildman reviews a range of ontologies of ultimacy (ungrounded nature, self-grounded nature, ground of being, personal being, symbiosis, non-moral dualism, and plural structures) in the light of what we know about the nature of the world. He endorses the characterization of the relation between law and chance in nature as 'law canalizing chance' and, using eight criteria, concludes that 'ground of being' and 'god-world symbiosis' ontologies of ultimacy are more compatible with laws canalizing chance than other ontologies.

Taking a global perspective, in Chapter Ten, John Bowker surveys beliefs from a number of religions that are relevant to the concepts of law and chance. He characterizes Karma as a moral law operating in the universe 'which is as certain as the law of gravity' (p. 184) and characterizes the actions of some deities as the explanation of apparently chancy events (because without such action the otherwise deterministic causal nature of the universe would not allow for chance). He goes on to highlight a common tension in a number of the world's religions, namely the tension between a single cause (God/Karma) and multiple causes (that allow for freedom). Another theme he examines is the concept of constraint. 'When world religions talk of creation (if they do), they are claiming that, in giving reasons why something, or everything, has happened, it is not possible to rule out God in the total specification of the constraints.' (p. 187)

And finally, in the Afterword, John Polkinghorne argues that our insights into the nature of physical laws (fragmented and imperfect as they are) point to the need for a deeper metaphysical explanation. 'The rationally transparent and beautiful principles of order already discerned as shaping cosmic process have a character that seems to call for further explanatory insight lying beyond that which science on its own can provide.' (p. 192) He suggests that complexity theory may be central to the advance of science, noting 'there may be undiscovered holistic laws of nature of a pattern forming kind' (p. 191), and that the concept of information may be as important to science in the next century as the concept of energy was during the last 150 years. Reflecting on the theme of the book, he observes: 'It is an important scientific insight that radical novelty ... [life, consciousness, and human self-consciousness] only emerges in regimes which can be thought of as existing "at the edge of chaos", domains where order and openness, chance and necessity, law and probability, intertwine.' (p. 190)

Having been somewhat critical earlier, I will end with praise. As someone with an active interest in the science-religion debate, I found this book interesting and valuable. It gave me insight into how a number of religious perspectives engage with the concepts of natural law and probability. And this has already borne fruit. For example, a number of contributors to the volume (e.g., Clayton, Gregerson and Wildman) make the observation that the regularity interpretation of laws of nature is arguably more compatible with theism than other interpretations. I had not appreciated this before and this observation has started me

thinking about the science-religion debate in new and fruitful ways. For example, given the compatibility between the regularity theory of laws of nature and the so-called 'block universe' theory of space-time, I am now prompted to consider the compatibility between the block universe and various theological perspectives. And to prompt such thinking is, surely, the very purpose of the book!